GENETIC ENGINEERING AND CHARACTERIZATION OF LYSR-TYPE TRANSCRIPTIONAL REGULATORS

By

Honghong Sun

RECOMMENDED:

Advisory Committee Chair

Department Head

APPROVED:

Dean, College of Science, Engineering, and Mathematics

Dean of the Graduate School

11-22-00

Date

GENETIC ENGINEERING AND CHARACTERIZATION OF LYSR-TYPE TRANSCRIPTIONAL REGULATORS

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By
Honghong Sun, M.S.

Fairbanks, Alaska
December 2000

Dedication

This work is dedicated to my parents,

to my daughter, Jenny Wei, and

to all of my teachers.

ABSTRACT

This thesis describes research aimed at understanding the structure and function of LysR-type transcriptional regulators. I studied two LysR-type proteins. One from the archaeon *Methanococcus jannaschii*, MJ-LysR. The other is from *Burkholderia cepacia*, DgdR. The MJ-LysR is the first putative LysR-type transcriptional regulator found in archaea. It is surprising that a prokaryotic transcriptional regulator is present in archaea, whose basal transcription machinery and RNA polymerase are more closely related to those of eukaryotes.

To elucidate the structure and function of MJ-LysR protein, the gene was subcloned and expressed in *E. coli*. The gene product was isolated and purified by heat treatment and size exclusion chromatography. An *in vitro* binding assay showed that the purified protein bound to the intergenic region between the *lysR* gene and its upstream gene specifically and selectively. The results also showed that the protein maintained its binding activity even at 94°C. The DNA footprinting data demonstrated a 30 bp protected region. Thus, this protein probably regulates expression of its own structural gene and perhaps the adjacent upstream gene.

DgdR protein from *Burkholderia cepacia* had been previously characterized (Allen-Daley et al. in preparation). The previous study showed that 2-methylalanine, the inducer for the DgdR regulated *dgdA* gene expression, but not D or L-alanine induced the conformational changes on DNA-protein complex. To further confirm this result, eleven amino acids with structures similar to 2-methylalanine were tested for their

ability on affecting the binding of the DgdR protein to its operator site. Among these amino acids tested, only 2-methylalanine, 1-aminocyclopentane-1-carboxylic acid, S-2-aminobutanoic acid, RS-isovaline, and 2-trifluoromethyl-2-aminobutanoic acid generated the measurable band shifting. D- or L-alanine, D- or L-norvaline, 2,2-diethyl glycine, and 2-trifluoromethylalanine did not cause any measurable change. It was concluded that both alkyl side chain size and hydrophobicity are important for the inducer recognition and binding in this protein.

To solve the problem in DgdR protein purification caused by low solubility of this protein, a *dgdR* fusion gene to *malE* gene was constructed. This fusion gene provides a useful tool to further study and crystallize the DgdR protein.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Appendices

Acknowledgments

I am grateful to many people for their generous help throughout my study at UAF. I would like to express my great gratitude to them sincerely.

First of all, I would like to thank my major research advisor, Dr. John Keller, for his encouragement and support throughout my course study, for his constructive suggestions and guidance for my research projects, especially for his confidence and encouragement during those hard days in my research. Special thanks to Dr. Lawrence Duffy for those good suggestions on protein purification and allowing me to be his teaching assistant for Biochemistry Lab for several semesters. I have learned a lot from it. I am also grateful to Dr. Joan Braddock and Dr. Thomas Clausen for their encouragement and assistance during my graduate study at UAF. I would like to thank Sheila Chapin in the Department of Chemistry and Biochemistry for what she has done to keep everything in order within the UAF system for me and for her warm smiles. I am also indebted to Marlys Schneider for her help with locating chemicals and equipment, and her wonderful chocolate cakes. Many thanks to my coworkers, Lilly Allen-Daley, Seetarn Woon, and Paoyu Chi, for their friendships and their help. I have had a lot of fun working with them. I would also like to acknowledge the financial assistance from the Department of Chemistry and Biochemistry.

Finally, I particularly thank my husband, Zhengyu Wei, for being supportive all the time and my parents for their love and inspiration.

Chapter 1

Literature Review:

Transcription, Transcription Regulation, and Related Issues in Bacteria, Eukaryotes, and Archaea

Introduction

Transcription is one of the important steps in gene expression. In order to provide cells or organisms the very protein(s) they need in certain situations or corresponding to certain external physiological signals, cells or organisms must transfer genetic information stored in the sequence of DNA to RNA. This process is called transcription, also known as RNA synthesis. In general, the overall process for transcription (RNA synthesis) on a double-stranded DNA molecule can be conceptually divided into three parts: initiation, elongation, and termination (Zubay 1995). In the initiation phase, the holoenzyme of RNA polymerase (RNAP) is formed and bound to the promoter region to form a "closed" complex. Then the "closed" complex is opened up by RNAP or other bound regulatory factors through melting about 12 base pairs and bringing in the necessary conformational changes in DNA-RNAP complex to expose the start point of transcription. This is so-called an "open" complex. After that, RNAP must escape from the promoter to start transcription. This step is known as promoter escape or promoter clearance. It involves a RNAP cycling reaction in which small abortively initiated RNA chains are produced; a RNA primer of sufficient stability is produced and is elongated into a productive RNA chain. In the elongation phase, the individual ribonucleotide

triphosphate basepairs with the template strand of the template DNA, followed by the stepwise formation of covalent bonds from one base to the next. Termination occurs when a specific sequence called a terminator is encountered by RNA polymerase. A double-strand RNA stem-and-loop structure is formed or a protein factor Rho is used to catalyze the termination process in bacteria. The process is not well understood in eukaryotes. The transcription ends up with break-up of the quaternary complex of DNA, RNA, RNA polymerase and the other possible factors.

Transcription can be regulated at any point during the process. However, the most efficient mode is at initiation phase. I will put my emphasis on this point in following discussions. The subtopics are listed as following:

1. Transcription and transcriptional regulation in the three domains of life
   1.1. Comparisons of basal transcriptional apparatus in the three domains
   1.2. Transcription regulation in bacteria
   1.3. General picture of eukaryotic RNA polymerase II Transcription and its regulation
   1.4. Transcription and its regulation in archaea.
2. DNA bending and transcriptional regulation
   2.1. Protein-induced DNA bending and bending angle
   2.2. Methods to detect protein-induced DNA bending
   2.3. DNA bending and bacterial transcription activation
3. The LysR family of bacterial transcriptional regulators (LTTRs)
   3.1. Background of the LysR family of transcriptional regulators

1. Transcription and transcriptional regulation in the three domains of life

Transcription occurs in all living organisms. However, the mechanism underlying its regulation is different in different organisms due to the structural differences of the transcription apparatus. For many years, it was thought that all life forms were divided into two domains: bacteria and eukaryotes. However, in 1977, C. Woese (Woese 1977) applied molecular biological techniques for classification purposes and suggested that archaea be separated from bacteria to form a third domain of life. Eukaryotes are distinguished from bacteria by possessing nuclei, extensive subcellular structures, compartmentalization, and more complicated RNA polymerase. Archaea are distinct from both bacteria and eukaryotes, and it is thought that they are more closely related to eukaryotes in molecular features although they lack nuclei, cytoskeleton, and organelles. Transcription and its control differ in the three domains. These will be addressed in the following sections.

1.1. Comparisons of basal transcriptional apparatus in the three domains

Before considering transcription regulation, the basal transcriptional apparatus must be described. The basal transcriptional apparatus is referred to a universal set of proteins that recognizes the core promoter and initiates transcription. It comprises RNA

polymerase and/or general transcription factors, such as TFIIA, TFIIB, TFIIF, TFIIH, TFIIE, TFIID in eukaryotes (Kornberg 1996).

In bacteria, RNA polymerase contains four subunits: $\alpha_2\beta\beta'$(core enzyme) and δ factor that is associated with the core enzyme more or less tightly. No more other factors are needed to start the basal transcription. Thus, the basal transcriptional apparatus is relatively simple in bacteria.

Eukaryotes have a more complex system in which three RNA polymerases synthesize different RNA molecules. RNA polymerase I is responsible for synthesis of pre-rRNAs. RNA polymerase III for pre-tRNAs and 5S RNAs. RNA polymerase II (RNA pol II) transcribes protein-coding genes. RNA pol II is an enzyme with 10-15 subunits. By itself, it cannot transcribe a gene, and requires the other general transcription factors, namely, TFIID (TBP-TAFs), TFIIA, TFIIB, TFIIE, TFIIH, and TFIIF. With the help of these factors, RNA polymerase II can access the promoter and form an initiation complex, followed by transcribing the gene.

Archaeal RNA polymerases contain 8-13 polypeptide subunits, with subunits sequences more similar to those of the eukaryotic RNA polymerase. One of the transcription factors found in archaea is the TATA-box binding protein (aTBP) whose structure is homologous to eukaryotic TBP (Marsh et al. 1994, Rowlands et al. 1994). The other archaeal transcription factor termed TFB is approximately 30% identical to TFIIB (Ouzounis & Sander 1992). Both *in vitro* and *in vivo* analyses have demonstrated that aTBP-TFB-RNA polymerase complex is sufficient to accurately start transcription in archaea (Hausner et al. 1996, Qureshi & Jackson 1996, Soppa 1999).

1.2. Transcription regulation in bacteria

1.2.1. Bacterial RNA polymerase

Transcription is the major control point of bacterial gene expression and DNA-dependent RNA polymerase (RNAP) is the central enzyme of the process. The core enzyme of RNAP from bacteria consists of four subunits: $\alpha_2\beta\beta'$. The precise function of each subunit is not known. However, the α-subunit has been found to perform at least three functions: (1) It initiates RNAP assembly by dimerizing into a platform with which the large β and β' subunits interact; (2) it participates in promoter recognition by sequence-specific protein-DNA interaction; and (3) it is the target of transcriptional activator proteins in transcription regulation (Ebright & Busby 1995). The catalytic site of RNA polymerase on the core enzyme is thought to be located on the β subunit. The β subunit also interacts with α subunit to form a stable assembly intermediate $\alpha_2\beta$ that recruits the β' subunit into the complex. The β' subunit is involved in interaction of RNAP with DNA at the promoter region as well (Brodolin et al. 2000). The α subunit connects these two large subunits into the core enzyme complex (Ishiama 1993). The core enzyme can catalytically extend the RNA chain by adding the corresponding ribonucleotide to the 3'-end of the growing chain. However, the core enzyme cannot recognize promoters and initiate specific RNA synthesis from double-stranded DNA templates by itself. Specific initiation of transcription in bacteria requires binding of a single polypeptide known as sigma (σ) factor to the core enzyme. This factor is involved in the recognition of promoters. To date, a total of six different sigma (σ) factors have been identified in *E. coli*. The σ^{70} with molecular weight of 70,000 Daltons

is the product of *rpoD* gene. Most regular gene expressions in growing cells are transcribed by this factor. Stress conditions, such as exposure of steady-state cells to heat shock, stimulate the expression of the *rpoH* gene whose product is σ^{32}. The σ^{24} is synthesized at extreme high temperature by *rpoE.* The *rpoN* gene encodes σ^{54}, required for transcription of genes that are controlled by the availability of nitrogen source (Helmann 1988). The *rpoF* gene encodes for σ^{28}, which is needed for transcription of a set of genes coding for the formation of flagella. The gene product of *rpoS* is σ^{38} that is essential for the expression of at least some stationary-phase-specific genes. Table 1-1 summarizes the functions of these factors.

The structure of *E. coli* RNA polymerase has been subjected to intense studies by several research groups. Although the individual RNAP subunits have not been crystallized. The high-resolution structural studies on the β, β', σ^{70}, and α subunits reveal that the individual subunits comprise important structural and functional subdomains (Blatter et al.1994, Darst 1998, Igarashi & Ishihama 1991, Jeon et al. 1995, Severinov et al. 1995, Wang et al. 1997). The α-subunit comprises two independent domains: the amino-terminal domain (αNTD, residues 8-235) and carboxyl-terminal domain (αCTD, residues 249-329). These two domains are connected by a flexible 14-residue linker. The αCTD is required for the interaction with UP elements located upstream of –35 element (Ross et al. 1993). It is the contact site for transcription activator as well (Ishihama 1993, Jeon et al.1995). It is composed of four short α helices, which form a compact fold as shown in Figure 1-1 (Jeon et al. 1995). The αNTD is necessary and sufficient for RNAP assembly and basal transcription. The X-

ray crystal structure of αNTD has been solved by Zhang and Darst (Zhang & Darst 1998). The αNTD monomer consists of two distinct, flexible linked domains, motif 1 and motif 2. Only motif 2 participates in dimerization. In the αNTD dimer, a long α-helix from one monomer interacts with the same helix structure from the other monomer to form an extensive hydrophobic core and bring these two monomers together (Zhang & Darst 1998). The principal determinants for dimerization of the α subunit are located within αNTD. A secondary determinant for dimerization is located with αCTD though it only provides a weak interaction. In overall structure, all the crucial determinants for interaction with the other RNAP subunits are located on the one face of the αNTD dimer; on the other side is the place for the carboxyl termini of the two αNTP monomers (Figure 1-1).

The crystal structure of a σ^{70} subunit fragment (residues 114-448) from *E. coli* RNAP has been solved by Malhotra et al.at Rockefeller University (1996). They found that the overall structure of this fragment (Figure 1-2) could be divided into three substructures that comprise two structural motifs. One motif consists of an antiparallel α-helical coiled-coil dimer with a helical bundle at one end. This motif is roughly repeated twice, giving rise to a "V"-shaped structure with pseudo-2-fold symmetry. The second motif comprises a small helical domain that is situated at the vertex of the V. The width of the V is about 50Å, and the distance separating the ends of the arms is about 75Å. In the perpendicular direction, the molecule is about 30 Å long. The residues 361-390, including most of the region known to interact with core RNAP face one side of the structure. On the opposite side, there are residues important for promoter

recognition, and the conserved aromatic residues involved in promoter melting are aligned along the solvent-exposed face of a single helix on the same side. So far, no structural information about β and β' subunits have been published. However, a cryo-electron microscope-derived density map of holo RNAP shows specific subunits and functional sites within the enzyme, and leads to a model for the organization of the subunits (Figure 1-3) (Darst et al. 1998).

1.2.2. Crucial DNA sequences in bacterial transcription regulation

Several specific DNA sequences are crucial for bacterial transcription regulation. They are core promoter element, UP element sequence, and upstream binding sites for regulatory proteins that includes operator sequences for repressor proteins and positive control elements for activator proteins. Promoters, as mentioned above, are the specific sequences where RNAP binds. The bacterial promoter contains two conserved hexamer sequences at the –10 and –35 positions relative to the transcription start site. They are called the –10 region/element and the -35 region/element. The DNA sequence separating these two regions is called the spacer. Usually, a promoter containing the consensus –10 element (TATAAT) and –35 element (TTGACA) is a very efficient promoter and specifies transcription by itself, no activators are needed. Typically, a 17-base spacer is between the –10 and –35 region. When the spacer is shorter or longer even by one base, the efficiency of a promoter can be reduced dramatically. However, very few natural promoters have the complete consensus hexamer sequences. Thus, ancillary factors might be needed in weaker promoters for the gene to be maximally

transcribed. The UP element is a DNA sequence rich in A and T. It is located at upstream of the –35 region. This element is recognized by the α-subunit of RNAP and is involved in stimulating transcription from the promoter. The other DNA sequences located upstream of the –35/-10 element that have an important influence on transcription are binding sites of regulatory proteins, in other words, the binding site for activators and repressors. For example, the activator protein binding on the upstream of the promoter makes protein-protein contact with RNAP and therefore provides additional stabilizing forces to start transcription or compensates for the weaker promoter. Figure 1-4 (Busby & Elbright 1994) shows the possible interactions among RNAP, promoter elements, and upstream regulatory sites.

1.2.3. Mechanism of transcriptional regulation in bacteria

The basic principle of gene regulation had been established in the mid-1960s by the pioneering work of Francois Jacob and Jacque Monod on the *E. coli lac* gene system. Their model has provided a tremendous stimulus for not only the gene regulation mechanism for the *lac* gene system but other genetic regulatory systems as well. In most bacterial systems, control of transcription is the dominant mode of gene regulation while the initiation phase for transcription is a major site for such a regulation. The fundamental units for gene regulation are: (1) promoter recognized by RNAP, (2) operator sequences or positive control elements recognized by repressor proteins or activator proteins respectively. Activators stimulate transcription by directly interacting with RNAP, at least 3 subunits, α, β', σ in RNAP can interact with activators. The

interaction between activators and subunits in RNAP can increase association of RNAP on the promoter through cooperative DNA binding, Or they can stimulate the activity of RNAP that is already bound to the promoter by facilitating the formation of an "open" complex and the clearance of the promoter (Struhl 1999, Hochschild 1998). Repressors function by blocking the activity of RNAP. This can occurre by simple occlusion of RNAP binding to the promoter or by generation of repressosome structures in which interactions between repressor molecules bound at distinct sites cause DNA loops (Struhl 1999), such as the inhibitory loop formed by lacI in *E. coli* (Lewis et al. 1996). Simple occlusion of RNAP binding can be achieved by two mechanisms. First, binding can be inhibited by physically blocking the promoter sequences. For example, in the *trp* operon, the *trp* promoter and *trp* operator regions overlap, the binding of repressor to the operator site prevents the binding of RNAP, thus inhibits the transcription (Zubay 1995). Secondly, distortion of the DNA molecule induced by binding of repressor can prevent the binding of RNAP. For example, in the *mer* operon, the binding of repressor in the absence of Hg introduced 25° twist in the promoter region and prevented the appropriate binding of RNAP (Ansari et al. 1992, Ansari et al.1995).

What actually underlies the phenomena here is two kinds of interactions, namely, protein-protein contact and protein-DNA interaction. The earlier studies showed that the changes in DNA structure due to the binding of protein to DNA molecules stimulated transcription activation. It is thought that changes in the DNA molecule upon binding of proteins were more important than protein-protein interaction for transcription activation (Perez-Martin & Espinosa 1993, 1991). The more recent studies on artificial

activators on the other hand demonstrated that protein-protein contacts also played an important role in transcription activation (Busby & Ebright 1994, Hochischild & Dove 1997, Hochischild & Dove 1998). The results from these research groups showed that an arbitrary interaction between a DNA-bound protein and RNA polymerase could activate transcription and if the activator makes contact with two different subunits of RNAP, the effect of these two interactions on transcription is synergistic. Therefore, a protein-protein contact model has been proposed (Dove & Hochschild 1998). In this model, any protein-protein interaction could, in principle, trigger transcriptional activation provided the relevant protein domains are fused to the α-NTD of RNAP and/or another DNA-binding protein. They further suggested that contact between a DNA-bound protein and any accessible surface of RNAP could activate transcription. These results lead to a conclusion that is opposite with the previous findings: the protein-protein contact instead of the changes in DNA structure upon the binding of a regulatory protein is a predominant force for bacterial transcription activation.

1.3. General picture of eukaryotic RNA polymerase II transcription and its regulation

As mentioned above, three different RNA polymerases are present in eukaryotes. Only RNA polymerase II (RNA pol II) is responsible for transcribing a protein-coding gene. To transcribe such a gene, RNA pol II, along with more than 20 proteins must be assembled at the promoter region (Buratowski 1994). The RNA pol II promoter is composed of core and regulatory (enhancer, silencer) regions. The core promoter region comprises a TATA box and transcription start site that is located about 30 base pairs

apart downstream. The regulatory region contains one or more sequences for interacting with DNA-binding regulatory proteins. Most of these sequences are located near upstream from the core element, but some might be at a great distance upstream from the core, some even at downstream from the core (Kornberg 1996). The RNA pol II transcription starts by binding of TFIID (complex of TATA box-binding protein, TBP and eight or more TBP-associated factors, TAFs) (Verrijzer & Tjian 1996, Hahn 1999) to the TATA box of a promoter. TFIIA contacts with TBP or TFIID complex and stabilizes the interaction of TFIID with the TATA box. This newly formed complex acts as a binding site for TFIIB that recruits RNA pol II and TFIIF into the complex. Subsequently, TFIIE and TFIIH associate with the initial complex. TFIIE recruits TFIIH into the pre-initiation complex, also modulates the helices and protein kinase activities of TFIIH while TFIIH provides helicase and protein kinase activities that are capable of phosphorylating the CTD of the RNA pol II and both enzyme activities are necessary for transcriptional initiation in eukaryotes (Buratowski 1994). The general picture of eukaryotic transcription is shown in Figure 1-5 (Roeder 1996).

In bacteria, the genomic DNA is associated with histone-like proteins, such as Hu in *E. coli* to form nucleoid structures. This nucleoid structure is unlikely to be a general inhibitory factor in transcription since RNAP can easily access the promoter region and start the basal transcription due to the inherent activity of promoters. Thus, the ground state of bacterial transcription is "nonrestrictive" (Struhl 1999). In contrast, in eukaryotes, the nuclear DNA was packed by histone proteins into nucleosome. The promoter region is buried in a nucleosome structure and is essentially inactive. Thus,

the ground state for transcriptional activity in eukaryotes is "restrictive" (Struhl 1999). In another words, the nucleosome structure represses all gene expressions except those whose transcription is brought about by specific positive regulatory mechanism (Kornberg 1999). That is the reason that most eukaryotic promoters are positively regulated while most bacterial promoters are negatively regulated.

In general, there are two classes of mechanisms for eukaryotic activators to stimulate transcription. Some activators start transcription by directly contacting RNA pol II. This case is analogous to the activation in bacteria (Ptashne & Gann 1997). For activators that do not contact RNA pol II directly, transcription activation is achieved through at least two steps: (1) relief of repression by activators and co-activators, and (2) function of mediator complex. Co-activator proteins were identified based on interaction with activators; furthermore, this group of proteins was discovered to acetylate the histone tails, a characteristic of transcribed chromatin. On the other hand, the histone deacetylase activities of co-repressors were found to silence gene expression. Just as co-activators can relieve repression by acetylating histone tails, co-repressor can establish repression by deacetylating histone tails. Another line of evidence might relate to the gene repression of deacetylated histone is DNA methylation (Huang et al. 1999). There is a long-standing correlation between DNA methylation and gene repression. It was reported lately by Ng (Ng & Bird 1999) and Jones (Jones et al. 1998) that the repression of gene expression by DNA methylation is through its effect on histone deacetylation. A protein called MeCP2 that binds to methylated DNA can recruit a multiprotein histone-deacetylase complex to deacetylate the histones and deacetylated histone makes the

transcription impossible. Therefore, by changing the structure of the histone molecule, the co-activators make it possible for RNA pol II to access the buried promoter and on the other hand, co-repressors make transcription impossible by converting histone molecules back to the transcription inactive form. However, activators and their associated co-activators failed to stimulate transcription in *in vitro* studies in certain cases (Flanagan 1992; Meisterernst 1991). This led to a conclusion that additional protein factors, termed mediators later, are required to trigger transcription *in vivo*. In yeast, this complex contains products of both novel and previously described genes, such as *srbs* and *gal11*, and binds to the CTD of RNA pol II to form holoenzyme. The function of mediators is important for activation and repression of transcription in eukaryotes. The possible model of mediator structure and function is shown in Figure 1-6. It is an important interface between activators and RNA pol II and transduces regulatory information from enhancers to promoters. However, many questions remain unclear about mediators. How do they influence RNA pol II and the initiation reaction? How do they contact activators and function to stimulate transcription?

Although the ground state of eukaryotic transcription is "restrictive" and chromatin structure modification itself is responsible for the restriction of eukaryotic gene expression, still, the restriction is not enough to completely silence gene expression. Consequently, eucaryotes have evolved specialized repression mechanisms, in which a DNA-binding protein recruits protein complexes to particular regions of the genome, and creates an extended domain of modified chromatin structure (Struhl 1999). For example, in yeast, the DNA-binding protein Rap1 and its associated co-repressors

recruit a complex of Sir proteins that can self-associate into a polymer-like structures to prevent the recruitment of a transcriptional initial complex by spreading along the chromosome (Hecht et al. 1996).

1.4. Transcription and its regulation in archaea

1.4.1. Archaeal RNA polymerase

The early observation about archaeal RNA polymerase (aRNA pol) of *Sulfolobus acidocaldarius* came from Zillig's lab (Zillig 1978, 1979). From the subunit component, which contains at least 10 subunits, it is apparent that aRNA pol is more similar to eucaryotic RNA polymerases (10-15 subunits) rather than bacterial RNAP ($\alpha_2\beta\beta'$). This conclusion was further demonstrated by serological cross-reactions (Huet et al. 1983). The recent findings showed that not only the subunit complexity but also the amino acid sequences in archaeal RNA pol are the more closely related to those of eukaryotic RNA polymerases (Langer et al.1995). Both large subunits and small subunits in aRNA pol are more closely homologous to those of the eukaryotic RNA pol II than the bacterial RNAP.

1.4.2. Archaeal promoter and basal transcription factors

The typical archaeal promoter contains (1) a TATA-like element, termed box A or the TATA box, located approximately 30 base pairs upstream of the transcription initiate site, (2) the purine-rich sequences immediately upstream of the TATA box known as transcription factor B recognition element (BRE), and (3) transcription initial

site, also called box B. The mutagenesis assays confirmed the key roles of the TATA box and the BRE sequences in archaeal transcription (Frey et al. 1990,Reiter et al.1990). Figure 1-7 shows the diagram of the general structure of an archaeal promoter.

To date, two kinds of transcription factors have been identified in archaea: an archaeal TATA-binding protein (aTBP) and transcription factor B protein (TFB) that recognizes the purine-rich BRE sequence upstream of TATA-box. aTBP is homologous to the eukaryotic TATA-box binding protein while the aTFB is homologous to the eukaryotic TFIIB. The homologs of eTBP have been found in a wide range of archaea and generally have 35%-40% identity with eTBPs (Bell & Jackson 1998a). The X-ray structure of *Pyrococcus woesei* TBP has been determined (DeDecker et al. 1996). The overall structure is similar to that of eTBP, possessing a saddle-like shape. The outer face of the saddle contains four α-helices and the inner DNA-binding face comprises an antiparallel β-sheet (DeDecker, 1996). One unique feature of aTBPs is that they possess a short carboxy-terminal tail with 6 to 10 amino acid residues long and rich in Asp and Glu, whose function, if any, remains unknown (Bell & Jackson 1998b). Another interesting aspect of aTBPs is that the electrostatic charge potential is distributed symmetrically in these proteins, unlike the same proteins in yeast and human, which have a highly asymmetric distribution with negative potential at the carboxyl-terminal loop, and positive potential at the amino-terminal loop (Bell et al. 1999b). This characteristic of aTBP might be relevant to orientation of transcription initial complex (Bell et al. 1999b). No TAFs (TBP-associated factors) have been reported in archaea. This may indicate that aTBP by itself can accurately recognize the TATA-box.

The second archaeal transcription factor, TFB, identified in 1992, is a homolog of eukaryotic TFIIB (eTFIIB)(Ouzounis & Sander 1992). It is about 30% identical to eTFIIB. The genes encoding TFB have now been isolated in variety of archaea (Creit et al.1993, Qureshi et al.1995). The structural organization of aTFB is very similar to that of eukaryotic TFII B. The aTFB has two domains. The structure of the amino-terminal domain of *Pyrococcus* TFB has been solved by NMR (Zhu 1996). The zinc-ribbon formed in this domain is similar to the zinc-ribbon in TFIIS, eukaryotic transcription elongation factor. The carboxyl-terminal domain contains an imperfect direct repeat. This domain can bind to a TBP-DNA complex to generate a stable ternary complex. The crystal structure of TFB with TBP and an oligonucleotide containing the TATA-box from an archaeal promoter (Kosa et al. 1994) indicates that this protein interacts with the carboxyl-terminal loop of TBP and with the DNA backbone on either side of the TATA-box as well.

1.4.3. Archaeal transcription and its regulation

The requirements for basal transcription in archaea have been subjected to intense studies and have been firmly characterized to date. Nevertheless, how regulation of gene expression is achieved remains largely unknown. Examining the complete genome sequences of various archaea reveals that the putative transcription regulators are more closely related to the bacterial transcription regulatory proteins. For example, the putative LysR-type transcriptional regulators were found in *Methanococcus jannaschii* (Bult et al. 1996) and *Archaeoglobus fulgidus* (Klenk et al. 1997). Four members of

ArsR family and nine members of AsnC family of bacterial transcription factors are present in *Archaeoglobus fulgidus* (Klenk et al. 1997). Recent work (Kyrpides & Ouzounis 1999) showed that among 280 predicted transcription factors or transcription-associated proteins in four archaeal genomes, 168 out of 280 have homologs only in bacteria, 51 only in eukaryotes, and 61 have homologs in both. This indicates that the bacterial transcription regulation mechanism might be applied in archaeal transcription regulation.

Although the mechanism underlying the archaeal transcription remains largely unknown, a few studies to date document mechanisms of transcription regulation in archaea. Nitrogen fixation in *Methanogenic* archaea has been used as a model system to study archaeal transcription regulation in Leigh and Zinder's Lab. Their results have shown that regulation of *nif* gene occurs by repression. Furthermore, repression is achieved by the specific binding of an unidentified factor to a palindromic element immediately downstream from the transcription initiation site (Cohen-Kupiec et al. 1997, Chien et al. 1998). The mechanism underlying it appears to resemble the bacterial model, with two subunits of a dimeric or tetrameric repressor protein binding to the operator palindromic sequences on the same face of the DNA helix. Another well-defined example of transcription regulation in archaea is the synthesis of gas vesicles in the extreme halophiles *Halobacterium salinarus* and *halöferam mediterranei* (Kruger 1998). A transcriptional activator containing a leucine-zipper motif, a DNA-binding motif characteristic in eukaryotes, was identified. When the gene for the major gas vesicle protein (*gvp*A) from *H. salinarus* was introduced into a non-gas vesicle

synthesizing species, *H. vocanii*, no *gvp*A transcription was detected. However, after *gvp*E gene was added in trans, *gvp*A transcription occurs. The sequence of *gvp*E reveals a leucine-zipper motif. The GvpE protein, the gene product of *gvp*E, binds to an appropriately spaced inverted repeat upstream of *gvp*A promoter to stimulate the transcription. This study demonstrates that a trans-acting transcriptional regulator in archaea resembles a class of eukaryotic DNA-binding transcription regulatory proteins (Kruger 1998). The most recent studies on archaeal transcription regulation come from S. Jackson's lab at the University of Cambridge. MDR1, a transcriptional regulator from *Archaeoglobus fulgidus,* is a homolog of bacterial metal-dependent transcription regulator, DtxR. MDR1 has been shown to bind to its own promoter specifically *in vitro*, moreover, this protein is specifically associated with its own promoter *in vivo* only under conditions where transcription is repressed. In addition, it negatively regulates transcription by preventing RNA polymerase recruitment in a reconstituted archaeal *in vitro* system. They concluded that the archaeal transcription system consists of bacterial-like regulators modulating the activity of an eukaryotic-like basal transcriptional apparatus (Bell & Jackson, 1999a).

As we know, in eukaryotes, one of the important aspects in transcription regulation lies in the modification of chromatin structure to facilitate assembly of the preinitiation complex. Whether the archaeal chromatin structure has the same effect on transcription regulation is still under investigation. The studies to date suggest that two different DNA compaction strategies be used in archaea. It will be mentioned again in a later section that archaea are subdivided into two highly divergent phylogenetic groups

according to16S rRNA sequence: the Euryarchaeota and the Crenarchaeota. The Euryarchaeota including the methanogens, halophiles, and some species of thermo/hyperthermophiles possess homologs of eukaryotic histones termed Hmf. The crystal structure of Hmf has been solved and the overall structure demonstrated to have high similarity to eucaryotic histone protein (Starich et al. 1997). To date, the reports concerning the effect of archaeal histones on the archaeal transcription have not been seen in publications. The Crenarchaeota, which include the sulfur-utilizing thermo/hyperthermophiles, on the other hand, do not appear to have histone homologs. They have small proteins known as Sac7d instead. This protein binds to the minor groove of DNA and induces a sharp kink to pack the DNA molecule. Whether or not this protein affects transcription remains unknown.

Another striking feature of the hyperthermophilic archaea worth mentioning here is its DNA topology. Unlike the DNA in eukaryotes and mesophilic bacteria with negatively supercoiled topology, the DNA in hyperthemophilic archaea is topologically relaxed or even positively supercoiled. How this feature affects archaeal transcription regulation is still unknown, but evidence indicates that the unique DNA topology in archaea contributes in some way to the thermostability of DNA molecules.

2. DNA bending and transcription regulation

An increasing body of evidence shows that transcriptional factors both from bacteria and eukaryotes bend DNA upon binding to their recognition sites. The DNA bending induced by transcription factors could regulate transcription in various ways. For

example, bending might bring transcription factors bound further upstream in contact with the general transcription apparatus, or loop DNA around to prevent the binding of RNA polymerase. Thus, protein-induced DNA bending is likely to be an important factor in transcription regulation. This section will only address the related issues of the role of protein-induced DNA bending in transcription although DNA bending could be induced by a variety of different means and bending also plays an important role in replication and recombination (Andrew 1990).

2.1. Protein-induced DNA bending and bending angle

DNA "bending" is defined as a deformation of DNA to a curvilinear shape caused by an external force. If a protein binds to DNA and bends the DNA fragment, the bending is known as protein-induced DNA bending. This should be distinguished from curved DNA caused by a run of adenine, A-track (Trifonov 1991). DNA bending angle (α) is defined as the angle by which a segment of the DNA departs from the linearity in Figure 1-8 (Thompson & Landy 1988).

2.2. Methods to detect protein-induced DNA bending.

Three different methods are currently used to detect protein-induced DNA bending. They are (1) Circular permutation analysis, (2) Phasing analysis, and (3) Cyclization analysis. Other methods, such as electron microscopy and measurement of rotational relaxation times of protein-DNA complexes are less frequently used (Van der Vliet & Verrijzer 1993).

2.2.1. Circular permutation analysis

The bent DNA can be detected by its anomalous migration in polyacrylamide gel (Wu & Crothers 1984). The mobility of protein-DNA complex is dependent on the end-to-end distance. Therefore, the position of protein-induced DNA bending determines the mobility of the complex. The bending located at the center of the DNA fragment leads to the most retarded mobility while bending at either end of the fragment yields the least retarded mobility. A DNA bending vector pBend2 was constructed for the purpose of circular permutation analysis (Kim et al. 1989). This is a vector can generate a large number of DNA fragments of identical length in which the protein-binding sequences are located at different positions in circular permutation manner. The protein-DNA complexes display different migration on a polyacrylamide gel due to the different locations of the bending (Figure 1-9). By plotting relative mobility against base pair position of the DNA, the bending is indicated and the bending center can be determined. An empirical relation between the degree of bending and the altered electrophoresis mobility on a polyacrylamide gel was used to calculate the bending angle (Thompson & Landy 1988). The formula is expressed as:

$$\frac{\mu_E}{\mu_M} = \cos\left(\alpha/2\right),$$

in which μ_E and μ_M stand for the mobility resulted from the fragment with the protein bound at end and center, α stands for the bending angle. This equation was modified to the more complicated form several years later (Zhou et al. 1993):

$$\mu_i / \mu_j = \frac{\left[1-2\left(x_i/L\right)(1-\cos\alpha)+2\left(x_i/L\right)^2(1-\cos\alpha)\right]^{0.5}}{\left[1-2\left(x_j/L\right)(1-\cos\alpha)+2\left(x_j/L\right)^2(1-\cos\alpha)\right]^{0.5}}$$

where μ_i and μ_j are the electrophoretic mobility of the protein-DNA complexes formed with the ith and jth DNA fragment, L is the total length of the DNA fragment, x_i and x_j are the distance from the left end of each fragment to the center of bending site on the fragment, α is the bending angle. Although this method is the most frequently used method to detect protein-induced DNA bending and calculate the bending angle, it is subject to several shortcomings. First, temperature, fragment length, and gel composition can strongly influence the electrophoretic mobility. Secondly, anomalous migration of the complex may also be caused by an aberrant shape of the protein itself instead of the protein-induced DNA bending (Van der Vlit & Verriijzer 1993).

2.2.2. Phasing analysis

Zinkel and Crothers introduced an elegant method to detect DNA bending direction and estimate the magnitude of the bending by comparing the electrophoretic mobility of isomers (Zinkel & Crothers 1987). These isomers have various helical phasing between the intrinsic bend caused by adenine-track (A-track) and protein-directed bends. The DNA fragment containing an A-track and a binding-site for a protein, separated by a set of linkers with different length was used in this analysis (Figure 1-10). Each different length of linker puts the protein-directed bends at a different phase relative to the

intrinsic curvature caused by the A-track. The mobility of DNA-protein complexes through a polyacrylamide gel exhibits striking shape-dependence for DNA molecules of identical molecular mass. The mobility of cis-isomer where two bends are in same phase would be minimal while the out-of-phase bends lead to maximal mobility due to reducing the overall bending (Zinkel & Crothers 1987). The resultant DNA fragments were ligated and resolved on a polyacrylamide gel. By comparing the retarded pattern of ligated bands, R_l (relative mobility) is determined. The relative mobility equals the ratio of apparent length to actual length. This ratio is a measure of DNA curvature (Strauss & Maher III, 1994). Calculation of the number of helical turns between A-track and protein-directed bends and examination of the gel pattern indicate the bending direction. The bending angle is solved by the following empirical equation:

$R_l-1=(\rho L^2-q)(\text{relative curvature})^2$,

in which R_l is relative mobility, L is the actual DNA length in base pair, ρ and q are constants, $5x10^{-5}$ and 0.5, respectively. The absolute curvature is the product of the relative curvature and the absolute curvature of the standard molecule (Crothers & Drak 1992, Zinkel & Crothers 1990).

This method is independent of the shape of the protein and can discriminate between protein-directed bends and other distortions of B-DNA structure because the mobility are dependent on the linker length instead of the direct end-to-end distance upon binding of proteins (Van der Vlit & Verriijzer 1993). It may be particularly useful for larger complex, for example, combinations of transcription factors and RNAP binding to the same fragment (Zinkle & Crothers 1991).

2.2.3. Cyclization analysis

The third independent way to monitor protein-directed bending is to measure the rate of cyclization of DNA fragments by DNA ligase. If a protein binds to a linear DNA fragment and bends it, the distance between the both ends of the DNA fragment will decrease. Therefore, with the presence of DNA ligase, the probability of ring closure will increase (Figure 1-11). By plotting percentage of linear fraction against time, DNA bending is indicated (Kotlarz et al. 1986). Although this method is relatively simple, reliable results are strongly dependent on protein-DNA ratio. When less protein below the saturation of the binding site is present, the method becomes less sensitive. On the other hand, too much protein may interfere ring closure, probably by non-specific binding to the ends.

2.3. DNA bending and bacterial transcription activation

Since the first observation that suggested that *E. coli* RNAP could introduce the bending to its bending site (Kuhnke et al., 1987), more and more evidence showed that DNA bends could be generated upon binding of transcription regulators. At the same time, intrinsic DNA curvatures were found within the bacterial promoter region and upstream of the –35 hexamer. Does the curved/bent DNA play an active role in cell functions, i.e. transcription? The last decade of research indicated that DNA not only carries genetic information for structural proteins and provides binding sites for regulatory proteins but also has intrinsic structural properties that play an active role in many cell functions (Perez-Martin et al. 1993, 1994).

How does the curved/bent DNA within promoter or upstream of promoter affect transcription activity or promoter activity? To answer this question, it is worth mentioning that transcriptional regulation is the major mode of bacterial gene regulation and transcription initiation is the major point of transcription regulation. The transcription initiation is a multistep process that starts at promoter recognition by RNAP, formation of a "close" complex and an "open" complex, concluding with promoter clearance, and synthesis of an oligonucleotide chain. As we know, the major determinants of promoter activity are the –35 region, -10 region, and spacer between them. RNAP recognizes the –35 region initially, subsequent contacts with –10 region, and results in formation of a "closed" complex. Dnase I and hydroxyl radical (OH*) protection analyses indicated that a bend was formed by simultaneous interaction of RNAP with –35 and –10 regions (Mecsas et al. 1991). This action realigns the –35 and –10 regions, thus placing stress on the spacer. The unfavored torsional strain in this state is thought to be the driving force to form an "open" complex. To complete this effect, not only an external force is needed but also the intrinsic potential bendability of DNA is required. The existence of the intrinsic curvatures upstream of the –35 hexamer coincides with the evidence that α subunit of RNAP contacts with the upstream AT-rich region in some promoters. The early model suggested that curved DNA sequences upstream might act as a docking region for RNAP, so that the concentration of RNAP was increased around the promoter region (Travers 1989, Plaskon & Wartel 1987). However, many results suggested that the actual scenario might be much more complicated. In most cases, the bends at the proximal site help the formation of a

"closed" complex by facilitating the initial wrapping of the DNA around the back of RNAP through direct DNA-protein interaction or through protein-protein interaction between the bound activator and RNAP. The bends at the distal site are also frequently involved with isomerization into an "open" complex. Furthermore, the further upstream contacts might help RNAP escape from promoter (promoter clearance) by springing the enzyme off the region. In other words, the melting domain at an "open" complex and promoter clearance appear to be the final results of releasing the torsional stress stored in DNA resulted by protein-induced DNA bending (Figure 1-12). On the other hand, DNA bending induced by protein facilitates protein-protein interaction. Bending brings regulatory proteins bound upstream closer to or even in contact with transcription basal apparatus bound at promoter region, otherwise, it would be impossible to interact with each other. The orientation and magnitude of the bend angle make it possible to select different interactions among proteins. This could explain transcriptional switch in which a regulatory protein can stimulate a gene expression from one promoter and at the same time repress another gene expression from another promoter. Although more and more evidence indicates the importance of protein-directed DNA bending and protein-protein interaction, the question remaining largely unknown is which interaction is actually more important in bacterial transcriptional regulation. In other words, is the bend just the consequence of protein binding or do the protein-protein interactions resulting by the bending play a key role in transcription activation? Do these two factors (DNA bending and protein-protein contact) have equal influence on transcription? Further studies are definitely needed to clarify this question.

For relationship between protein-induced DNA bending and bacterial transcription repression, it is generally believed that the negative regulatory proteins inhibit transcription by physically blocking the binding site of RNAP. Recent studies challenge this concept by finding: (1) RNAP binding site is not mutually occupied by repressor in certain promoters, (2) Some repressors have multiple binding sites located at a certain distance from the RNAP binding sites. In both cases, the higher-order structure of DNA-protein or protein-protein might prevent the access of RNAP. The binding of repressor might cause disorientation/distortion in DNA structure or form a loop that maintain promoter in an unsuitable comformation for transcription.

3. The LysR family of bacterial transcriptional regulators (LTTRs)

3.1. Background of the LysR family of transcriptional regulators

3.1.1. Common features of the LysR family of transcriptional regulators

The LysR family of transcriptional regulators was first identified by Henikoff et al. (1988). Since then the family has grown from the original 9 members to over 100 members now. The members in this family regulate metabolisms in a wide range of procaryotic species. Recently, the LysR-type proteins were found in hyperthermophilic archaea (Bult et al. 1996; Klenk et al. 1997; Deckert 1998). The protein in this family regulates transcription of one or more metabolic genes and its own structural gene. Several general characteristics are shared among these proteins (Schell 1993):

(1) They all have a highly conserved N-terminal amino acid sequences that is predicted to adopt helix-turn-helix DNA-binding motif.

(2) The size of the protein from this family ranges from 274-324 amino acids.

(3) Each is divergently transcribed from a promoter that is very close to or overlaps with a promoter of its cognated target gene.

(4) Most LTTRs stimulate the transcription of their regulated genes while repressing the expression of their own structural genes.

(5) Most LTTRs are coinducer-responsive transcriptional regulators.

DNA alignments among LTTRs have shown that the greatest amino acid sequence identity is within the first 66 N-terminal amino acids. The other less identical regions are located at C-terminal, position 95-173 and position 196-206. The secondary structure prediction also suggested that a three- or four-domain model that coincides with the sequencing alignment. In this model, the N-terminal highly conserved amino acids form a helix-turn-helix structure that is the most frequently used DNA-binding motif in procaryotes. The less conserved C-terminal is also thought to be important for DNA interactions. It may play a role in the interaction between LTTR subunits and/or protein-DNA interactions occuring during transcription regulation. The conserved amino acids at the middle of peptides may be involved in recognition of coinducer and function as a binding pocket, at same time, allowing the conformational changes caused by binding of coinducer.

3.1.2. Structural and functional indications for other LysR-type transcriptional activator from three-dimensional crystal structure of one LTTR: CysB

Cysteine biosynthesis in Gram-negative bacteria involves the products of a total of 22 *cys* genes, 14 of which are coordinately regulated and collectively referred to as the cysteine regulon. CysB protein controls the expression of the cysteine regulon. It is a member of the LysR family of transcriptional regulators. The crystal structure of a dimeric fragment of *Klebsiella aerogenes* CysB comprising of residues 88-324 has been recently solved and refined to 1.8 Å resolution (Tyrrell et al. 1997). It provides the first direct structural insight into three-dimensional structure of LysR family proteins.

The overall structure of monomer adopts ellipsoid-shape. It consists of two similar α/β domains, domain I and domain II, connected by two crossover regions of residues 163-166 and residues 266-269. As shown in Figure 1-13, domain I consists of five β strands and four α helices that are made up of two regions of polypeptide chain: residues 88-162, and residues 270-292. The β-sheet formed by five β-strands (β_B β_A β_C β_J β_D) was extended by the sixth β-strand (β_K) formed by residues 322-324 at the C-terminus. Three of four helices are located in the loop regions between each β-strands, the fourth one is located after β_j. Domain II is made up of a single stretch of polypeptide chain consisting of residues 166-265. It contains five-strand β sheet interspersed with three α-helices. The order of the strands is $\beta_I\beta_E\beta_H\beta_F\beta_G$. The three α-helices lie after β_E, β_F, and β_G. A cleft is formed between the two domains with four threonines and one glutamine (Thr100, Thr102, Thr149, Thr201, and Gln103) on its inner surface. This is

thought to be the binding site for anti-inducer, sulfate anion. In the dimer, domain I from the first monomer comes together with domain II from the second monomer in a side-by-side manner through H-bonding between β_B and β_G. As a result, an 11-stranded β sheet is formed. Upon the formation of dimer, two clefts in monomers are brought together to form an internal channel that is thought to be the binding site for inducer, *N*-acetylserine.

Unfortunately, the possible insight into the mode of DNA binding was limited due to the lack of the first 87 N-terminal residues. However, the previous mutation analyses have shown the great abolishment of DNA binding activity by mutating the N-terminal residues. Combined with the secondary structure predictions, the N-terminus is considered as a region that recognizes and binds to the promoter.

3.2. Possible existence of LysR-type proteins that negatively regulate their cognate target gene expressions

Although almost all of the studied transcriptional regulators in the LysR family regulate their cognate target gene expressions positively, the studies on DgdR, a member of the LysR family in Keller's Lab did show that this protein was a negative transcriptional regulator (Keller et al. 1990). The DgdR protein, the gene product of the *dgdR*, and the product of its regulated *dgdA* control metabolism of 2,2-dialkylglycines, such as 2-methylalanine in *Burkholderia cepacia*. The promoter of the divergent *dgdR* gene overlaps with that of the *dgdA* gene. The *dgdA* gene encodes a vitamin B-6-dependent dialkylglycine decarboxylase. The expression of the *dgdA* gene is under

the control of the DgdR protein. A partial-deletion of the *dgdR* gene and the *xylE* reporter gene insertion experiments demonstrated the negative control of the *dgdA* gene by the DgdR protein (Allen-Daley et al. in preparation). The Dnase I footprinting assay further revealed the shrink of the DgdR binding site upon binding of the co-inducer, 2-methylalanine. This indicated that DgdR protein physically blocks the binding site of RNAP in the absence of the co-inducer while freeing the binding site in the presence of the co-inducer (Allen-Daley et al. in preparation). The results from the DgdR protein indicate the existence of negative transcriptional regulators in the LysR family.

3.3. Transcription control/DNA-binding proteins in thermophiles/hyperthermophiles

The complete genome sequences of *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, and *Aquifex aeolicus* reveal the existence of homologs of the LysR protein in archaea. A BLAST search of Genbank database using the amino acid sequence of any of LysR proteins generates over 100 open reading frames (ORFs) with a similarity greater than 25%. Among these hits, five are from hyperthermophilic archaea: one from *M. jannaschii*, one from *A. fulgidus*, two from *A. aeolicus*, and one from *Cyanidium caldarium* (Keller 1999). The existence of homologs of the bacterial transcriptional regulatory proteins in archaea, instead of eukaryotic transcription general factors might imply the bacterial gene regulation mode in archaea.

3.3.1. Definition and classification of hyperthermophilic archaea

Hyperthermophiles consist of a group of microorganisms, whose optimum growth temperatures are above 80°C and maximum growth temperatures are around 110°C. The group includes some species from archaea and a number of bacteria (Stetter et al. 1990). Archaea are now considered to be a monophyletic group. It was not recognized as a taxonomic group until late 1970s (Woese et al. 1977). Archaea can be conveniently divided into three phenotypes according to their inhabiting environments: the thermo/hyperthermophiles, methanogens, and halophiles (Woese 1987). However, according to 16S rRNA sequence comparisons, archaea can be subdivided into two highly divergent phylogenetic groups (Figure 1-14): the Euryarchaeota and the Crenarchaeota (Woese 1987, 1990). The Euryarchaeota includes the methanogens, halobacteria, and some species of hyperthermophiles such as *Thermococcus* (Gary et al. 1997). Later, this group was found to possess homologs of eukaryotic histones. These archaeal histones compact DNA into nucleosome-like structure, in which the packing mode is highly similar to that of eukaryotes (Pereira et al. 1971, Starich et al. 1996). The Crenarchaeota studied to date do not have eukaryotic histone homologs. This category consists of many the sulfur-utilizing hyperthermophiles, such as *Thermoproteus* and *Pyrodictium* (Gary et al. 1997). They have small proteins known as Sac 7d to pack their DNAs. The X-ray crystal structure reveals that this protein binds and bends DNA in the minor groove and forms a higher-ordered structure. This result indicated the second type of archaeal chromatin structure that might be representative of the Crenarchaeota (Robinson et al. 1998).

3.3.2. Possible mechanisms of thermostability of biomacromolecules

In order to survive extreme environments, such as high temperature, all the components of archaeal cells must be heat-resistant. These components include lipids, proteins, and nucleic acids. However, cell components in mesophilic bacteria are quite heat-sensitive. Although the molecular basis for heat adaptation of biomacromolecules in thermophiles is still under investigation, some studies show the structural differences that might contribute to the thermal stability of these biological molecules. One study showed that archaeal membranes contain ether lipid derived from diphytanyl-glycerol or its dimer. These two chemicals exhibit strong resistance against high temperature and an acidic pH (Kates 1992), in contrast, the mesophilic counterparts have ester lipid.

The thermal stability of DNA is thought to be predominantly contributed by a reverse gyrase. This unique type I DNA topoisomerase was found in all archaeal hyperthermophilic species, but not in any of mesophilic or moderately thermophilic organisms (optimal growth temperature below 60°C). It might improve the thermostability of DNA molecules by introducing positively supercoiled turns in DNA structure at the expense of ATP (Forterre 1996, Stetter 1999). Another possible contribution to DNA stability under high temperature might come from archaeal histone proteins. These proteins are homologous to the eukaryotic histone proteins, but tolerate much higher temperatures than their mesophilic counterparts. They might provide further thermostability by wrapping around the DNA molecules (Li et al. 2000, Stetter 1999). *In vitro* experiments showed that addition of histones to purified DNA increased its melting point dramatically (Woese et al. 1991).

Protein heat adaption has been subjected to intense studies. The comparisons of amino acid sequences showed a very high identity between hyperthermophiles and their mesophilic counterparts. They both use the same amino acids as building blocks and adopt the same tertiary structures. What underlies the shift of denature temperature? The studies suggested that protein thermostability is unlikely to be achieved by big changes in protein structures due to the small net free energy of stability. It is more likely to be achieved by addition of small changes at different locations in the protein molecules (Pace et al. 1996). It seems apparent that thermostability is a result of a combination of several interactions, such as increased numbers of hydrogen-bonding, increased ionic interaction and forming salt bridge network, the burying as much as possible the hydrophobic residues inside the protein and increasing the hydrophobic interactions, and reducing chain flexibility. Different proteins might use different combinations (Ladenstein & Antranikian 1998). The recent research on adenylate kinases (ADK) from several species of *Methanococcus* showed that both terminal regions of the ADKs made a major contribution to the their thermostability by the hydrophobic interaction of residues contributed by both ends. Further mutation analysis of these hydrophobic residues demonstrated the role of a hydrophobic core in determining protein stability (Haney et al. 1999). Besides, these internal factors coming from the macromolecules themselves, extrinsic factors, such as K^+, cyclic 2,3-diphosphoglycerate (cDPG), can stabilize the native state of the protein at high temperature. High concentrations of these chemicals were demonstrated to increase protein thermostability *in vitro* (Hensel 1996). The heat-shock proteins may also be

essential for growth at high temperature (Phipps et al. 1991). These chaperones might be involved in the repairing and refolding of proteins that have undergone damages by heat. The underlying mechanism is still unclear. Deeper insights into the stabilizing mechanism of hyperthermophilic proteins are still under investigation. These insights are most likely to be achieved by comparison of 3-D structures with their mesophilic counterparts.

Project objectives

It has been two decades since archaea was first announced to be the third kingdom of life. Since then, it has been of particular research interest. Archaea constitutes a wide range of organisms that grow in a wide variety of extreme environments. The research on archaea not only sheds light on life evolutionary history but also provides unique opportunities to study thermostability of proteins, DNA and other macromolecules, mechanisms of protein folding, and other aspects of gene expression. The complete genome sequence of *Methanococcus jannaschii* reveals a putative LysR-type transcriptional regulator. The goal of our research here is to confirm the existence of a LysR-type transcriptional regulator in this archaeon, study the binding activity of this protein, identify its binding site on DNA, and finally to try to elucidate the transcription mechanism in this archaeon.

Table 1-1. Sigma factors in *E. coli.* (Adapted from Ishihama 1993)

Sigmal Factors	Genes	-35 region Recognized	–10 region Recognized	Transcribed genes
σ^{70}	*rpoD*	TTGACA	TATAAT	Regular genes
σ^{32}	*rpoH*	CTTGAA	CCCCAT-TA	Heat-shock genes
σ^{24}	*rpoE*	CTGGCACNNN	NNTTGCA	Extreme heat-shock genes
σ^{54}	*rpoN*	GAACTT	TCTGA	Nitrogen-regulated genes
σ^{28}	*rpoF*	TAAA	GCCGATAA	Flagella-chemotaxis genes
σ^{38}	*rpoS*	UNKNOWN	UNKNOWN	Stationary phase-specific genes

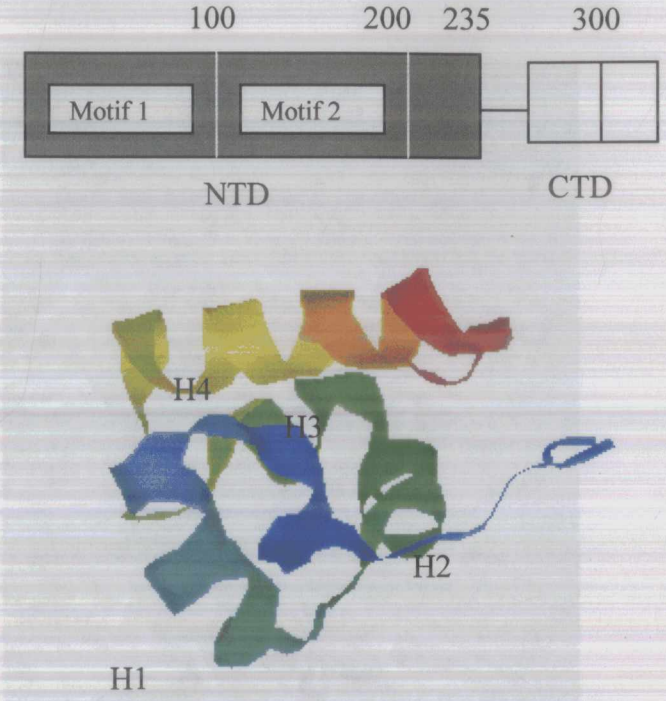


Figure 1-1a. (top) Schematic diagram showing the domain structure of the *E. coli* RNAP α-subunit. (bottom) The 3-D structure of C-terminal domain of the *E. coli* RNAP α-subunit (Darst et al.1998, Jeon et al. 1995, Malhotra & Severinova 1996).

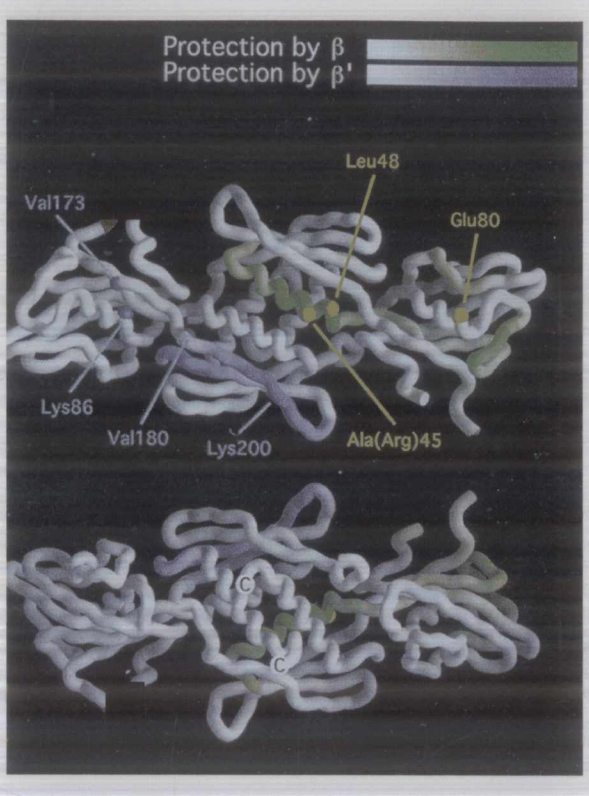


Figure 1-1b. The 3-D structure of N-terminal domain of α subunit. of *E. coli* RNAP(top) The region protected by β, β'are colored green and magenta, respectively. For clarity, the region protected by β are shown on one monomer while the region protected by β' on the other. The amino acids shown in yellow and light blue are mutations that abolish the β or β' binding. (bottom) The opposite view of the top picture. The carboxyl termini of the two αNTD monomers are labeled by C (Darst et al. 1998).

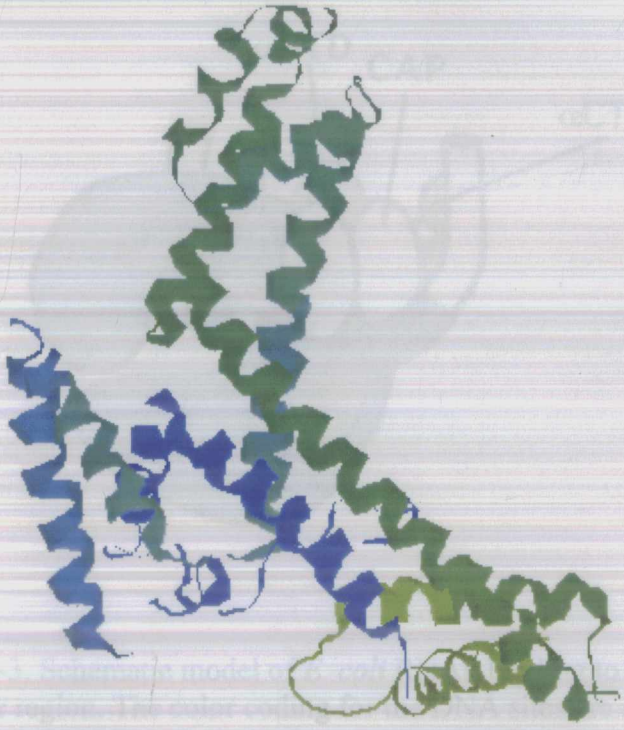


Figure 1-2. The 3-D structure of σ-subunit fragment (114-448 bp) of *E. coli* RNAP (Malhotra & Severinova 1994).

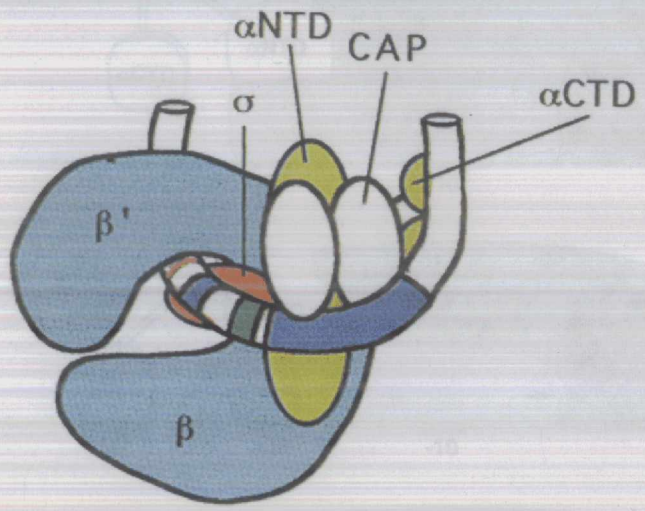


Figure 1-3. Schematic model of *E. coli* RNAP binding to the promoter region. The color coding for the DNA sites are as follows: red line, transcription start site; magenta, -10 region of promoter; light green, -35 region of promoter; violet, CAP binding site. The color coding for proteins are labeled (Darst et al. 1998).

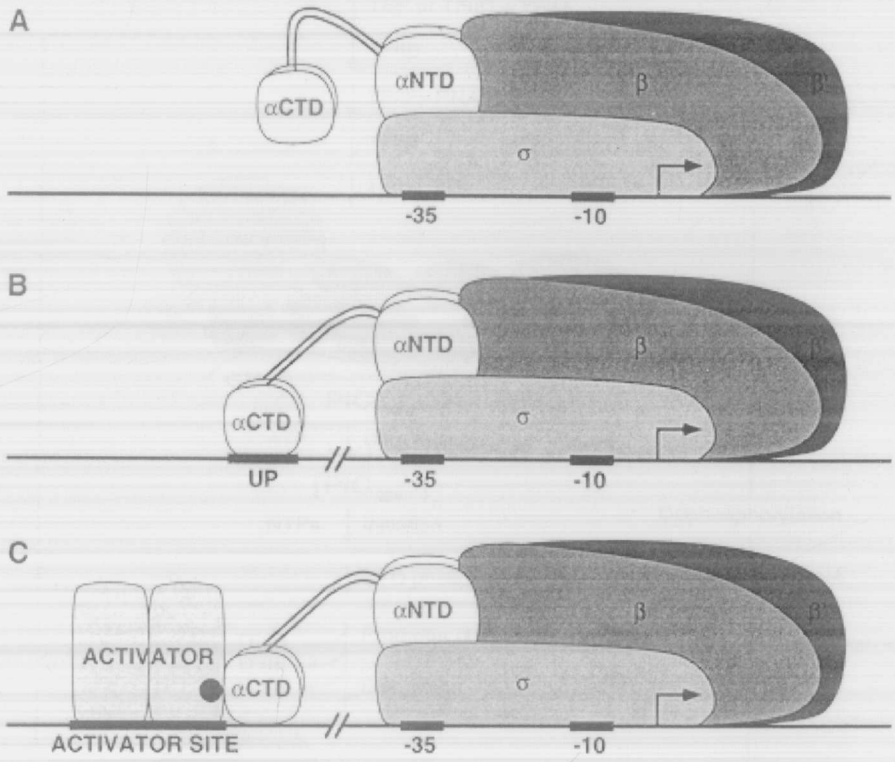


Figure 1-4. Prokaryotic transcription complexes. (A) Simple promoter, αCTD makes no specific contacts. (B) αCTD contacts with UP element in promoter containing UP element. (C) αCTD contacts with the activator bound upstream (Busby & Ebright 1994).

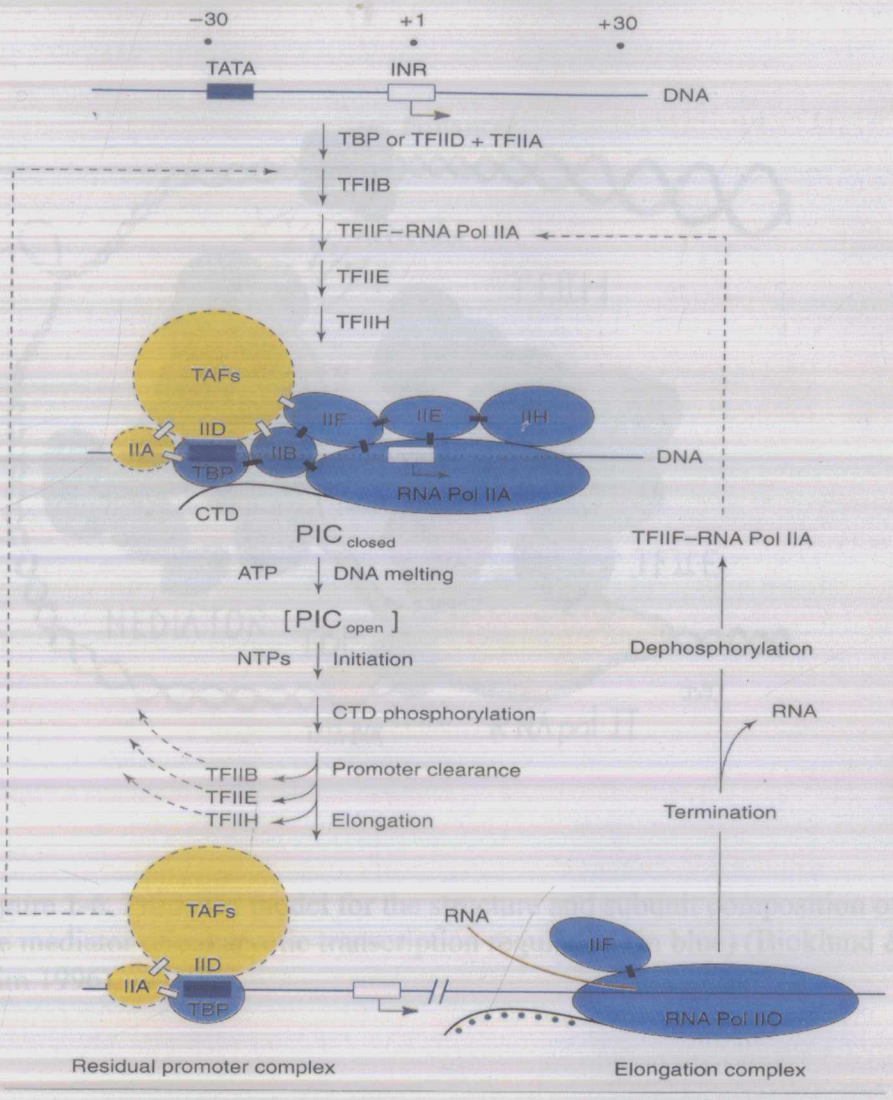


Figure 1-5. General picture of eukaryotic transcription (Roeder 1996)

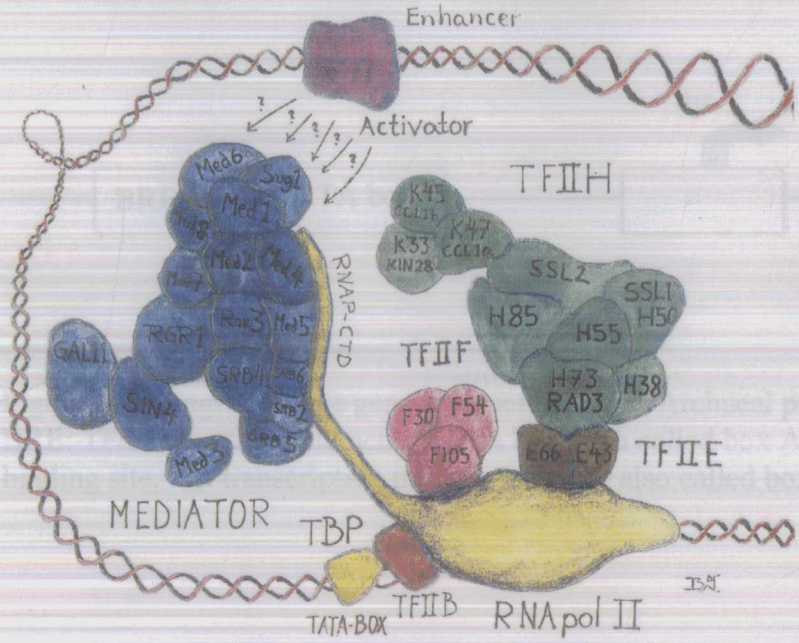


Figure 1-6. Proposed model for the structure and subunit composition of the mediator of eukaryotic transcription regulation (in blue) (Bjoklund & Kim 1996).

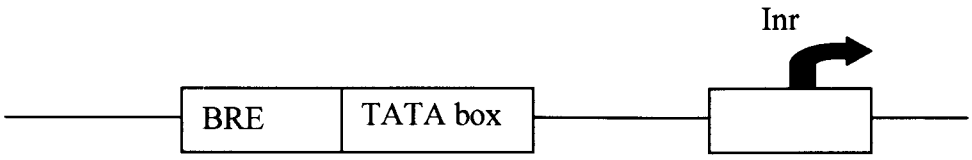


Figure 1-7. Diagram of the general structure of an Archaeal promoter. BRE: TFB recognition element; TATA box: also called box A, aTBP binding site. Inr: transcription initiator element, also called box B.

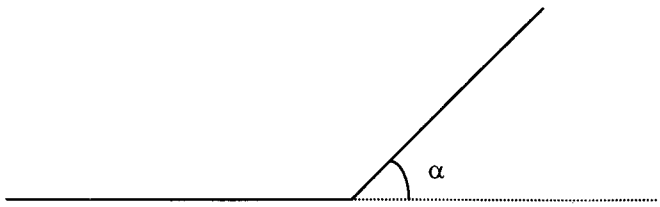


Figure 1-8. Diagram of bending angle induced by a protein.

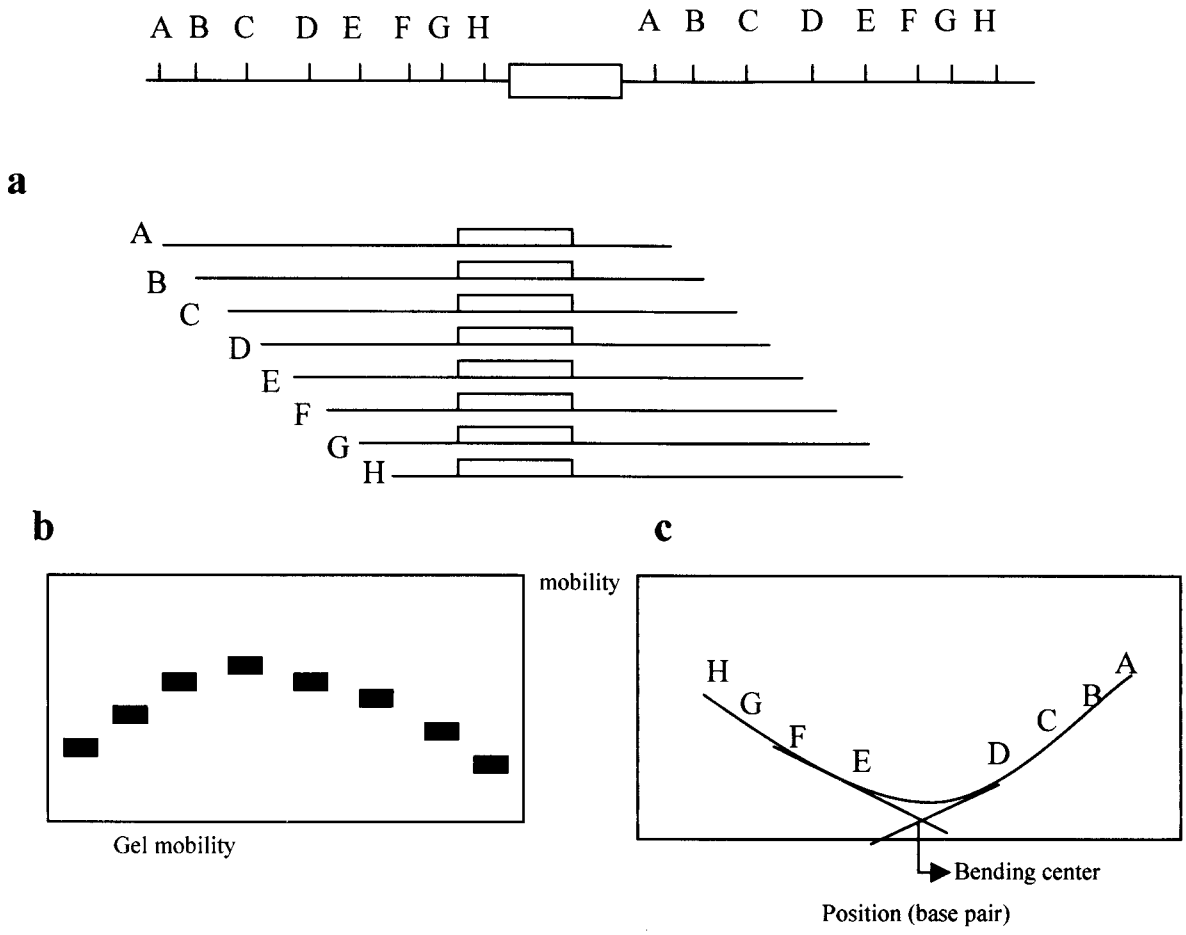


Figure 1-9. Circular permutation assay. A-H: Restriction enzyme cutting sites. (a) DNA fragments with binding sites at different position in permutation manner. (b) Band pattern on a polyacrylamide gel. (c) Plotting mobility against base pair position to find bending center.

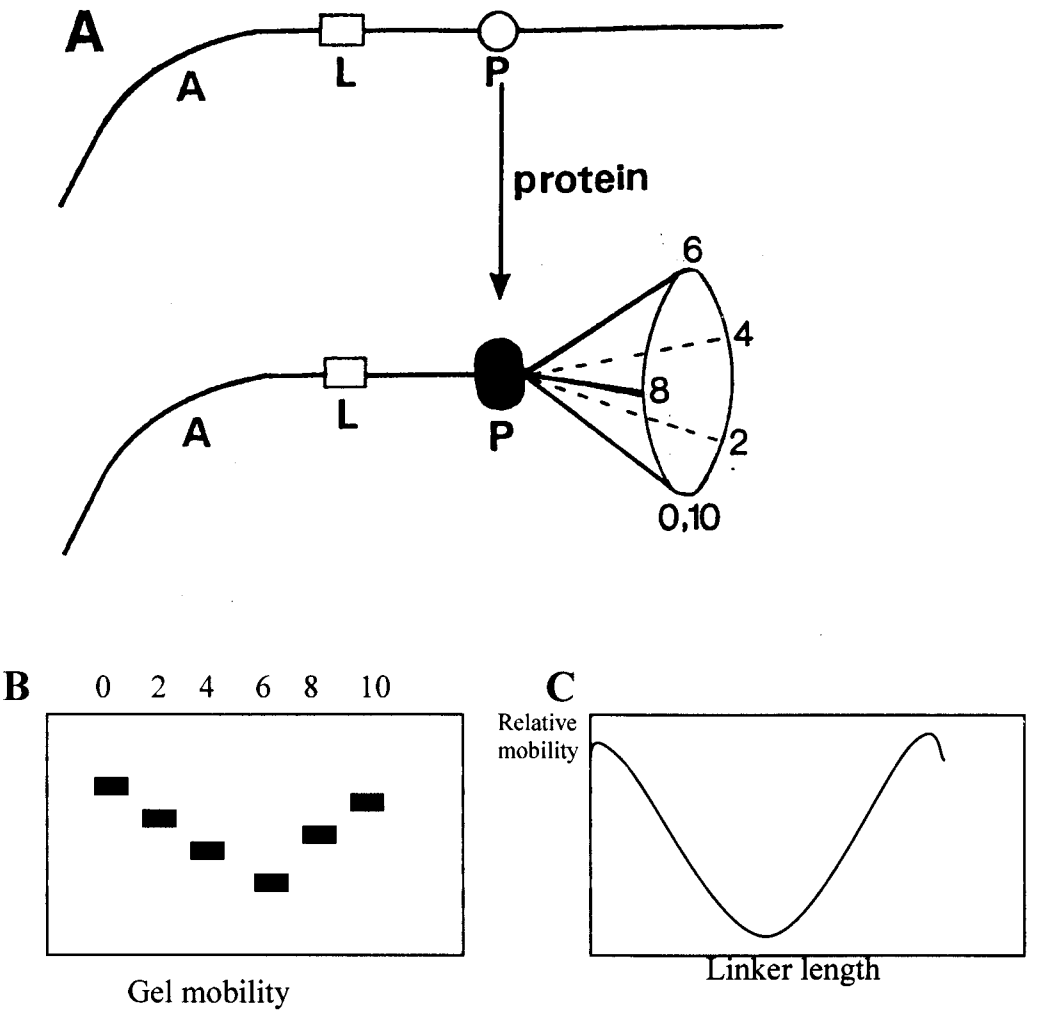


Figure 1-10. Phasing analysis. (A)Diagram of the assay. The A-track directed curvature A and a protein binding site (P) are connected by a set of linkers with different length, thus generate a set of isomers with the protein-directed bending angle in different phase with the curvature A. (B) Gel mobility pattern on polyacrylamide gel. (C) The plotting of relative mobility against linker length indicates the bending .

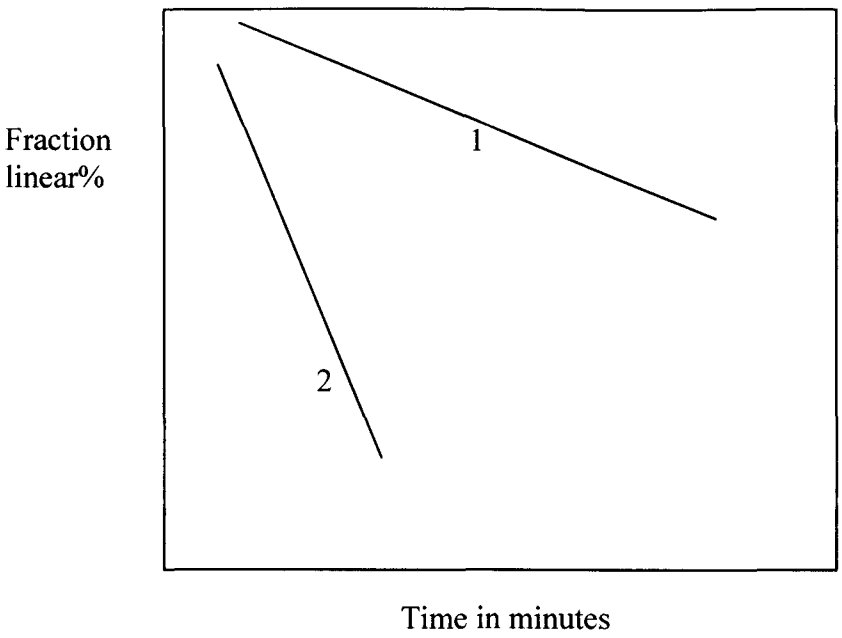


Figure 1-11. Cyclization analysis. Line 1 fraction linear percentage against time with the absence of the DNA bending protein; line 2 with the presence of the protein.

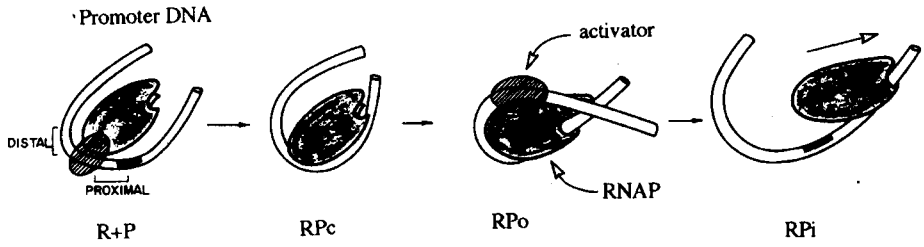


Figure 1-12. General model for bacterial promoters responsive to DNA bending. R+P: RNAP interacts with promoter; RPc: the "closed" initial complex formed by RNAP and promoter; Rpo: the "open" initial complex formed by RNAP, activator, and promoter; Rpi: RNAP escapes from promoter and starts transcription (Perez-Martin et al. 1994).

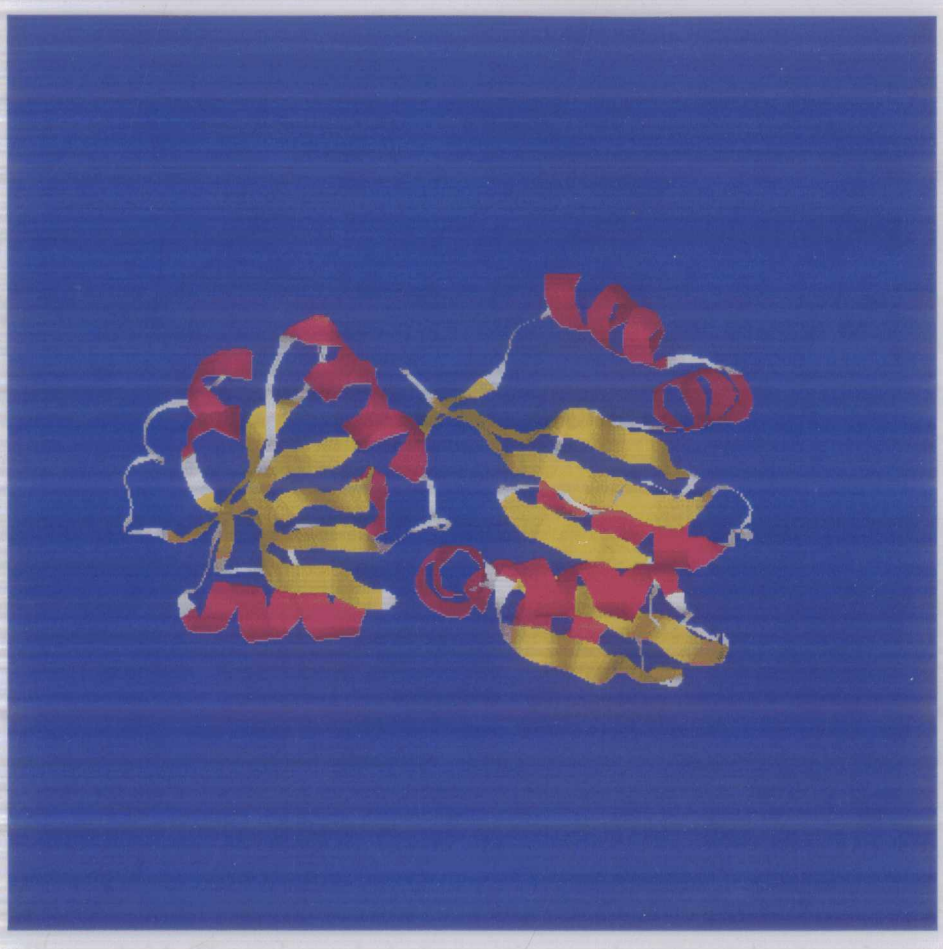


Figure 1-13. The secondary structure of the CysB(88-324) monomer (Tyrrell et al. 1997).

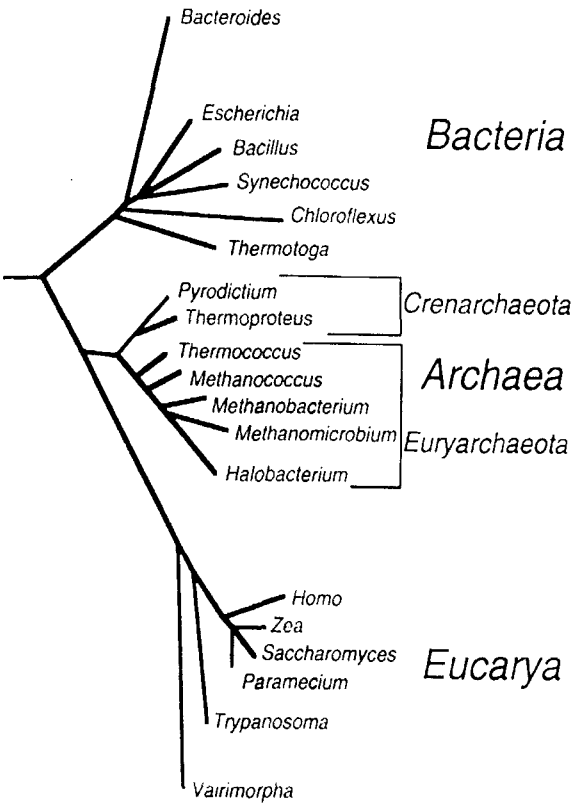


Figure 1-14. The universal phylogenetic tree showing three domains of life (Olsen & Woese 1997).

References:

Adams, M.W.W. and Kelly, R.M. 1994. Thermostability and thermoactivity of enzymes from hyperthermolphilic Archaea. *Bioorg. Med. Chem.* 2(7): 659-667.

Allen-Daley, E., O'Brien, M.L., Bray-Hall, S.T, Stapleton, R.D., Chi, P., Sun, H., and Keller, J.W. Characterization of the *Burkholderia (Pseudomonas) cepacia* dgdR gene encoding a LysR-type negative regulator of 2,2-dialkylglycine metabolism. (in preparation).

Anasri, A.Z., Chael, M.L. and O'Halloran, T.V. 1992. Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* 355: 87-89.

Ansari, A.Z., Bradner, J.E., and O'Halloran, T.V. 1995. DNA-bend modulation in a repressor-to-activator switching mechanism. *Nature* 374: 371-375.

Bell, S.D. and Jackson, S.P. 1998a. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* 6: 222-228.

Bell, S.D. and Jackson, S.P. 1998b. Transcription in Archaea. *Cold Spring Harbor Symp. Quant. Biol.*. LXIII, 41-51.

Bell, S.D., Cairns, S.S., Robson, R.L., and Jackson, S.P. 1999a. Transcriptional regulation of an Archaeal operon *in vivo* and *in vitro. Mol. Cell.* 4: 971-982.

Bell, S.D., Kosa, P.L.,Sigler, P.B., and Jackson, S.P. 1999b. Orientation of the transcription preinitiation complex in Archaea. *Proc. Natl. Acad. Sci. USA.* 96: 13662-13667.

Bjorklund, S. and Kim, Y. 1996. Mediator of transcriptional regulation. *Trends Biochem.* 21: 335-337.

Bjorklund, S., Almouzni, G., Davidson, I., Nightgale, K.P., and Weiss, K. 1999. Global transcription regulators of Eukaryotes. *Cell* 96: 759-767.

Brodolin, K., Mustaev, A., Severinov, K., and Nikiforov, V. 2000. Identification of RNA polymerase beta subunit segment contacting the melted region of the *lacUV5* promoter. *J. Biol. Chem.* 275(4): 3661-3666.

Bult,C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, K.D., et al. 1996. Complete genome sequence of the Methanogenic Archaeon, *Methanococcus jannaschii. Science* 273: 1058-1073.

Buratowski, S. 1994. The basics of basal transcription by RNA polymerase II. *Cell* 77: 1-3.

Busby, S.and Ebright, R.H. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743-748.

Chien, Y., Helmann, J.D., and Zinder, S.H. 1998. Interactions between the promoter regions of nitrogenases Structural Gene (*nif*HDK2) and DNA-binding proteins from N_2- and ammonium-grown cells of the archaeon *Methanococcus barkeri* 227. *J. Bacteriol.* 180: 2723-2728.

Choy, H.E. and Adhya, S. 1993. RNA polymerase idling and clearance in *gal* promoters: use of supercoiled minicircle DNA template made *in vivo. Proc. Natl. Acad. Sci. USA.* 90: 472-476.

Cohen-Kupiec, R., Blank, C., and Leigh, J.A. 1997. Transcriptional regulation in archaea: *In vivo* demonstration of a repressor binding site in a methanogen. *Proc. Natl. Acad. Sci. USA.* 94: 1316-1320.

Cowan, D.A. 1992. Biotechnology of the Archaea. *Trends Biotech.* 10: 315-323.

Creti, R., Londei, P., and Cammarano, P. 1993. Complete nucleotide sequence of an archaeal (*Pyrococcus woesei*) gene encoding a homologue of eukaryoteic transcription factor IIB. *Nucleic Acids Res.* 21: 2942.

Crothers, D.M. and Drak, J. 1992. Global features of DNA structure by comparative gel electrophoresis. *Methods Enzymol.* 212: 47-71.

Crothers, D.M., Drak, J., Kahn, J., and Levene, S.D. 1992. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* 212: 3-29.

Darst, S.A., Polyakov, A., and Zhang, G. 1998. Structural studies of *Escherichia coli* RNA polymerase. *Cold Spring Harbaor Symp. Quant. Biol.* LXIII: 269-276.

Deckert,G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J. and Swanson, R.V. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392: 353-358.

DeDecker, B.S., O'Brien, R., Fleming, P.J., Geiger, J.H.m Jackson, S.P., and Sigler, P.B. 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* 264: 1072-1084.

Diekmann, S. 1992. Analyzing DNA curvature in polyacrylamide gels. *Methods Enzymol.* 212: 30-47.

Dove, S.L. and Hochschild, A. 1998. Use of artificial activators to define a role for protein-protein and protein-DNA contacts in transcriptional activation. *Cold Spring Harbor Symp. Quant. Biol.* LXHIII: 173-180.

Dove, S.L., Joung, J.K.,and Hoshschild, A. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* 386: 627-630.

Elbright, R.H. and Busby, S. 1995. The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr. Opin. Gen. Deve.*5: 197-203.

Flanagan, P.M., Kelleher III, R.J., Sayre, M.H., and Kornberg, R.D. 1991. A mediator required for activation of RNA polymerase II transcription *in vitro. Nature* 350: 436-438.

Forterre, P., Bergerat, A., and Lopezgarcia, P. 1996. The unique DNA topology and DNA topoismerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* 18: 237-248.

Forterre, P., Bergerat, A., and Lopez-Garcia, P. 1996. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* 18: 237-248.

Frey, G., Thomm, M., Brudigam, B., Gohl, H.P., and Hausner, W. 1990. An archaeobacterial cell free transcription system the expression of transfer RNA genes from *Methanococcus vannielii* is mediated by a transcription factor. *Nucleic Acids Res.* 18: 136-142.

Geanacopoulos, M., Vasmatzis, G., Lewis, D.E.A., Roy, S., Lee, B., and Adhya, S. 1999. GalR mutants defective in repressosome formation. *Gene & Deve.* 13: 1251-1262.

Goldstein, M. and Doi, R.H. 1995. Prokaryotic promoters in biotechnology. *Biotechnol. Annu. Rev.* 1: 105-128.

Gralla, J.D. 1990. Promoter recognition and mRNA initiation by *Escherichia coli* σ^{70}. *Methods Enzymol.* 185: 37-54.

Hahn, S. 1998. The roles of TAFs in RNA polymerase II transcription. *Cell* 95: 579-582.

Haney, P.J., Stees, M., and Konisky, J. 1999. Analysis of thermal stabilizing interactions in mesophilic and thermophilic Adenylate Kinase from the genus *Methanococcus. J. Biol. Chem.* 274(40): 28453-28458.

Hausner, W. and Thomm, M. 1993. Purification and characterization of a general transcription factor, aTFB, from the archaeon *Methanococcus thermolithotrophicus. J. Biol. Chem.* 268: 24047-240.

Hausner, W., Frey, G., and Thomm, M. 1991. Control regions of an archaeal gene a TATA box and an initiator element promoter cell free transcription of the tRNA(Val) gene of *Methanococcus vannielii. J. Mol. Biol.* 222: 495-508.

Hausner, W., Wettach, J., Hethke, C., and Thomm, M. 1996. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an Archaeal RNA polymerase. *J. Biol. Chem.* 271: 30144-30148.

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. 1996. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383: 92-96.

Helmann, J.D. and Chamberlin, M.J. 1988. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57: 839-872.

Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. 1988. A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. USA.* 85: 6602-6606.

Hochschild, A. and Dove, S.L. 1998. Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* 92: 597-600.

Huang, L., Guan, R.J., and Pardee, A.B. Evolution of transcriptional control from Prokaryotic beginnings to Eukaryotic complexities. *Critical Rev. in Eukaryo. Gene Expr.* 9(3&4): 175-182.

Hudepohl, U., Reiter, W.D., and Zillig, W. 1990. *In vitro* transcription of two rRNA genes of the archaebacterium *Sulfolobus sp. B12* indicates a factor requirement for specific initiation. *Proc. Natl. Acad. Sci. USA* 87: 5851-5855.

Ishihama, A. 1981. Subunit assembly of *Escherichia coli* RNA polymerase. *Adv. Biophys.* 11: 1-35.

Ishihama, A. 1988. Promoter selectivity of Prokaryotic RNA polymerase. *Trends Genet.* 4: 282-286.

Ishihama, A. 1990. Molecular assembly and functional modulation of *Escherichia coli* RNA polymerase. *Adv. Biophys.* 26: 19-31.

Ishihama, A. 1993. Protein-protein communication within the transcription apparatus. *J. Bacteriol.* 175(9): 2483-2489.

Jaenicke, R. 1996. Glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: strategies of protein stabilization. *FEMS Microbiol. Rev.* 18: 215-224.

Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. 1995. Solution structure of the activator contact domain of the RNA polymeras α subunit. *Science* 270: 1495-1497.

Jones, P.L., Veenstra, G.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P.1998. Methylated DNA and MeCP2 recruit histome deacetylase to repress transcription. *Nat. Genet.* 19: 187-191.

Kates, M. 1992. The Archaebacteria: Biochemistry and biotechnology. p51-72. Portland Press, London and Chapel Hill.

Keller, J.W., Baurick, K.B., Rutt, G.C., O'Malley, M.V., Sonafrank, N.L., Reynolds, R.A., Ebbesson, L.O., and Vajdos, F.F. 1990. *Pseudomonas cepacia* 2,2-dialkylglycine decarboxylase. *J. Biol. Chem.* 265(10): 5531-5539.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, K.D., Richardson, D.L.,Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Sill, S., Kirkness, E.F., Dougherty, B.A., Mckenney, K., Adams, M.D., Loftus, B., Venter, J.C., et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate reducing archaeon *Archaeoglobus fulgidus. Nature* 390: 364-370.

Kornberg, R.D. 1996. RNA polymerase II transcription control. *Trends Biochem.* 21: 325-326.

Kornberg, R.D. 1999. Eukaryotic transcriptional control. *Trends Biochem.* 24: M46-M49.

Kosa, P.F., Ghosh, G., DeDecker, B.S., and Sigler, P.B. 1997. The 2.1 Å crystal structure of an archaeal preinitiation complex:TATA box binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. USA* 94: 6042-604.

Kruger, K., Hermann, T., Armbruster, V., and Pfeifer, F. 1998. The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. *J. Mol. Biol.* 279: 761-771.

Kyrpides, N.C. and Ouzounis, C.A. 1999. Transcription in archaea. *Proc. Natl. Acad. Sci. USA.* 96: 8545-8550.

Ladenstein, R. and Antranikian, G. 1998. Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. engi.* 61: 37-85.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. 1998. New core promoter element in RNA polymerase II dependent transcription: sequence specific DNA binding by trancription factor IIB. *Gene Deve.* 12: 34-44.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. 1995. Transcription in archaea similarity to that in eucaya. *Porc. Natl. Acad. Sci. USA* 92: 5768-5772.

Leigh, J.A. 1999. Transcriptional regulation in archaea. *Curr. Opin. Microbiol.* 2: 131-134.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. 1996. Crystal structure of the lactose complexes with DNA and inducer. *Science* 271: 1247-1254.

Li, W., Shriver, J.W., and Reeve, J.N. 2000. Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. *J Bacteriol.* 182(3): 812-817.

Malhotra, A. and Severinova, E. 1996. Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* 87: 127-136.

Mecsas, J., Cowing, D.W., and Gross, G.A. 1991. Development of RNA polymerase-promoter contacts during open complex formation. *J. Mol. Biol.* 220: 585-597

Meisterernst, M., Roy, A.L., Lieu, H.M., and Roeder, R.G. 1991. Activation of Class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* 66: 981-993.

Miller, A., Wood, D., Ebright, R.H., and Rothman-Denes, L.B. 1997. RNA polymerase β' subunit: a target of DNA binding-independent activation. *Science* 275: 1655-1657.

Ng, H. and Bird, A. 1999. DNA methylation and chromation modification. *Curr. Opin. Genet. & Deve.* 9: 158-163.

Niu, W., Kim, Y., Tau, G, Heyduk, T., and Ebright, R.H. 1996. Transcription activation at class II CAP-dependnet promoters: two interactions between CAP and RNA polymerase. *Cell* 87: 1123-1134.

Olsen, G.J. and Woese, C.R. 1997. Archaeal genomics: an overview. *Cell* 89: 991-994.

Oost. J.V., Ciaramella, M., Moracci, M., Piasni, F.M., Rossi, M., and DeVos, W. M. Molecular biology of hyperthermophilic archaea. *Adv. Biochem. Engi/biotech.* 61: 88-115.

Ouzounis, C. and Sander, C. 1992. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell* 71:189-190.

Pace, C.N., Shirley, B.A., Mcnutt, M., and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J.* 10: 75-83.

Palmer, J.R., and Daniels, C.J. 1994. A transcriptional reporter for *in vivo* promoter analysis in the archaeon *Haloferax volcanii. Appl. Environ. Microbiol.* 60: 3867-3869.

Palmer, J.R., and Daniels, C.J. 1995. *In vivo* definition of an archaeal promoter. *J. Bacteriol.* 177: 1844-1849.

Pereira, S.L., Grayling, R.A., Lurz, R.,and Reeve, J.N. 1997. Archaeal nucleosomes. *Proc. Natl. Acad. Sci. USA* 94: 12633-12637.

Perez-Martin, J. and de Lorenzo, V. 1997. Clues and consequences of DNA bending in trancription. *Annu. Rev. Microbiol.* 51: 593-628.

Perez-Martin, J. and Espinoas, M. 1991. The RepA repressor can act as a transcriptional activator by inducing DNA bends. *EMBO J.* 10(6): 1375-1382.

Perez-Martin, J. and Espinosa, M. 1993. Protein-induced bending as a transcriptional switch. *Science* 260: 805-807.

Perez-Martin, J., Rojo, F., and de Lorenzo, V. 1994. Promoters responsive to DNA bending: a common theme in Prokaryotic gene expression. *Microbiol. Rev.* 58(2): 268-290.

Phipps, B.M., Hoffmann, A., Stetter, K.O., and Baumeister, W. 1991. A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaeobacteria. *EMBO J.* 10(7): 1711-1722.

Plaskon, R.R. and Wartel, R.M. 1987. Sequence distributions associated with DNA curvature are formed upstream of strong *E .coli* promoter. *Nucleic acids Res.*15: 785-797.

Pineiro, S., Olekhnovich, I., and Gussin, G.N. 1997. DNA bending by the TrpI protein of *Pseudomonas aeruginosa. J. Bacteriol.* 179(17): 5407-5413.

Qureshi, S.A., Khoo, B., Baumann, P., and Jackson, S.P. 1995.Molecular cloning of the transcription factor TFIIB homolog from *Sulfobobus shibatae. Proc. Natl. Acad. Sci. USA* 92: 6077-6081.

Reeve, J. N., Sandman, K., and Daniels, C.J. Archaeal histones, nucleosomes, and transcription initiation. *Cell* 89: 999-1002.

Reiter, W.D., Hudepohl, U., and Zillig, W. 1990. Mutational analysis of an archaeobacterial promoter: Essential role of a TATA-box for transcription efficiency and start site selection *in vitro*. *Proc. Natl. Acad. Sci. USA* 87: 9509-9513.

Rice, D.W., Yip, K.S.P., Stillman, T.J., Britton, K.J., Fuentes, A., Connerton, I., Pasquo, A., Scandurra, R., and Engel, P.C. 1996. Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol. Rev.* 18: 105-117.

Robinson, H., Gao, Y., McCrary, B.S., Edmondson, S.P., Shiver, J.W., and Wang, A.H. 1998. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* 392: 202-205.

Roeder, R. and Pfeifer, F. 1996. Influence of salt on the transcription of the gas vesicle genes of haloferax mediterranei and identification of the endogenous transctriptional activator gene. *Microbiol.* 142: 1715-1723.

Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem.* 21: 327-335.

Ross, W., Gosink, K., Salomon, J., Igarashi K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262:1407-1413.

Rowlands, T., Baumann, P., and Jackson, S.P. 1994. The TATA-binding protein a general transcription factor in eukaryotes and archaeobacteria. *Science* 264: 1326-1328.

Schell, M.A. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* 47: 597-626.

Schnetz, K. and Wang, J.C. 1996. Silencing of the *Escherichia coli bgl* promoter: effects of template supercoiling and cell extracts on promoter activity *in vitro*. *Nucleic Acids Res.* 24(12): 2422-2428.

Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E.N., Roberts, J.W., Chait, B.T. and Darst, S.A. 1996. Domain organization of the *Escherichia coli* RNA polymerase σ^{70} subunit. *J. Mol. Biol.* 263: 637-647.

Soppa, J. 1999. Transcription initiation in archaea: facts, factors and future aspects. *Mol. Microbiol.* 31(5): 1295-1305.

Starich, M.R., Sandman, K., Reeve, J.N., and Summers, M.F. 1996. NMR structure of Hmfb from hyperthermophile, *Methanothermus fervidus,* confirms that this archaeal protein is a histone. *J. Mol. Biol.* 255: 187

Stetter, K.O. 1999. Extremophiles and their adaptation to hot enviroments. *FEBS Lett.* 452: 22-25.

Stetter, K.O., Fiala, G., Huber, G., and Segerer, A. 1990. *FEMS Microbiol. Rev.* 75: 117-124.

Strauss, J.K. and Maher III, L.J. 1994. DNA bending by asymmetric phosphate neutralization. *Science* 260: 1829-1834.

Struhl, K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98: 1-4.

Thomm, M. 1996. Archaeal transcription factors and their roles in transcription initiation. *FEMS Microbiol. Rev.* 18: 159-171.

Thompson, J.F. and Landy, A. 1988. Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucleic Acids Res.* 16(20): 9687-9705.

Travers, A. and Muskhelishvili, G. 1998. DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J. Mol. Biol.* 279: 1027-1043.

Travers, A.A. 1990. Why bent DNA? *Cell* 60: 177-180.

Travers, A.A. 1989. DNA conformation and protein binding. *Annu. Rev. Biochem.* 58: 427-452.

Trifonov, E.N. 1991. DNA in profile. *Trends Biochem.* 16: 467-470.

Tyrrell, R., Verschueren, K.HG., Dodson, E.J., Murshudov, G.N., Addy, C., and Wilkinson, A. 1997. The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Struc.* 5(8): 1017-1032.

Van der Oost, J., Ciaramella, M., Moracci, M., Pisani, F.M., Rossi, M., and De Vos, W.M. 1998. Molecular biology of hyperthermophilic archaea. *Adv. Biochem. Engi.* 61: 87-115.

Van der Vliet, P.C. and Verrijzer, C.P. 1993. Bending of DNA by transcription factors. *BioEssays* 15(1): 25-32.

Verrijzer, C.P. and Tjian, R. 1996. TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem.* 21: 338-342.

Verrijzer, C.P., Van Oosterhout, J.A.W.M., Van Weperen, W.W., and Van der Vliet, P.C. 1991. POU proteins bend DNA via the POU-specific domain. *EMBO J.* 10(10): 3007-3014.

Wang, Y., Severinov, K., Loizos, N., Fenyo, D., Heyduk, E., Chait, B.T. and Darst, S.A. 1997. Determinants for *Escherichia coli* RNA polymerase assembly within the β subunit. *J. Mol. Biol.* 270: 648-662.

Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* 51(2): 221-271.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Prod. Natl. Acad. Sci. USA* 87: 4576-4579.

Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* 74: 5088-5090.

Xiao, L. and Honig, B. 1999. Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* 289: 1435-1444.

Zhou, Y., Zhang, X., and Ebright, R.H. 1993. Identification of the activation region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proc. Natl. Acad. Sci. USA* 90: 6081-6085.

Zhu, W.L., Zeng, Q.D., Colangelo, C.M., Lewis, L.M., Summers, M.F., and Scott, R.A. 1996. The N terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat. Struct. Biol.* 3: 122-124

Zillig, W., Prangishvilli, D., Schleper, C., Elferink, M., Holz, I., Alvers, S., Janekovic, D., and Gotz, D. 1996. Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. *FEMS Microbiol. Rev.* 18: 225-236.

Zillig, W., Stetter, K.O., and Janekovic, D. 1979. DNA-dependent RNA polymerade from the Archaebacterium *Sulfolobus acidocaldaruus*. *Eur. J. Biochem.*96: 597-604.

Zillig, W., Stetter, K.O., and Tobien.M. 1978. DNA dependent RNA polymerase from *Halobacterum halobium*. *Eur. J. Biochem.* 91: 193-199.

Zinkel, S.S. and Crother.D.M. 1987. DNA bend direction by phase sensitive detection. *Nature* 328:178-181.

Zinkel, S.S. and Crothers, D.M. 1990. Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein. *Biopoly.* 29: 29-38.

Zinkel, S.S. and Crothers, D.M. 1991. Catabolite activator protein-induced DNA bending in transcription initiation. *J. Mol. Biol.* 219: 201-215.

Zinkel, S.S. and Crothers, M. 1991. Catabolite Activator protein-induced DNA bending in transcription initiation. *J. Mol. Biol.* 219: 201-215.

Chapter 2

Cloning, Expression, and Characterization of a LysR-type Transcriptional Regulator from the Archaeon *Methanococcus jannaschii*

Abstract

The complete genome sequence of the archaeon *Methanococcus jannaschii* reveals the presence of a gene encoding a putative LysR-type transcriptional regulator (Bult et al. 1996). This 296-amino acid protein is encoded by the *MJ0300* gene whose promoter overlaps with that of a nearby divergent upstream gene of unknown function. The *MJ0300* gene was subcloned into the bacterial expression vector pET-5b (plasmid vector) and was transformed by competent *E. coli* BL2(DE3) pLysS cells. The gene expression has been successfully induced by addition of 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) in the growth medium. The MJ0300 gene product was isolated and purified by heat-treatment and size exclusion chromatography. The purified recombinant protein is about 75 KD judged by size exclusion chromatography, implying a dimer of two 36.8 KD subunits. The purified protein binds to the regulatory region of its nearby upstream gene specifically and selectively in an *in vitro* binding assay. In addition, the binding activity of the purified protein at different temperatures was tested by a gel mobility shift assay. The result showed that the protein maintained its binding activity when heated to 94°C. DNA footprinting data demonstrated a 30 bp region protected by this protein on the DNA. This result is consistent with current knowledge about the size and location of binding sites of other LysR-type

transcriptional regulators in bacteria. These results might indicate the possible existence of a bacterial transcription regulation mechanism in this archaeon.

Introduction

Since the LysR family of transcriptional regulators was first identified by Henikoff (1988), the family has grown from 9 members to over 100 members. Proteins in this family have been found to be involved in a wide range of metabolic pathways in bacteria. However, despite the efforts of many researchers, the structures of these proteins remain unclear. Amino acid alignments among these proteins have shown several highly or moderately conserved regions. The first 60 amino acids at the N-terminus among the family members are highly conserved and are predicted to carry the helix-turn-helix structure that is the most frequently used DNA recognition and binding motif in bacteria. The less conserved regions including the regions from residues 90 to 180 and residues196 to 206 are thought to be involved in coinducer recognition/response, and the C-terminal residues 220-250 are also important for DNA interaction or coinducer recognition (Schell 1993). The crystal structure of dimeric CysB fragment missing the first 87 residues, a LysR-type protein from *Klebsiella aerogenes,* has been solved recently (Tyrrell et al. 1996). The overall shape of the dimeric CysB fragment is ellipsoidal. The two monomers are brought together in a side-by-side manner. The inner channel formed by Thr100, Thr102, Thr149, Trp166, and Thr202 is likely to be a binding site for the inducer, *N*-acetylserine. This result is consistent with the predictions of domain organizations derived from DNA sequence

alignments. The other common characteristics shared by the proteins in this family are listed as follows. All the members have a similar size ranging from 276-324 amino acid residues. The LysR-type proteins divergently regulate the expressions of their cognate genes. The promoters of these LysR-type proteins overlap with that of their regulated genes. Most of the proteins in this family activate their regulated gene expressions in the presence of inducers while repressing their own gene expressions. The binding sites of the LysR-type proteins almost invariantly contain a so-called LysR motif: T-N_{11}-A within an imperfect, inverted repeat, usually located around 50-60bp upstream of the transcription start site of their regulated genes (Byerly et al.1991, Schell & Poser 1989, Goethals et al.1992).

Genes encoding putative homologs of LysR-type transcriptional regulators have been found in the genomes of several archaea. The first Archaeon found to carry the homolog of the LysR-type transcriptional regulator was *Methanococcus jannaschii. M. jannaschii* was first isolated from a sediment sample collected from the sea floor at the base of a 2600-meter-deep "white smoker" chimney which is located on the East Pacific rise (Jones et al. 1983). This archaeon is a strict anaerobe and methanogen. It grows at pressures of more than 200 atm and over a temperature range of 48°C to 94°C, with an optimum growth temperature near 85°C. The putative LysR-type transcriptional regulator was identified by a computer-based sequencing analysis of *M. jannaschii* (Bult et al. 1996). This is quite surprising because the results from previous research have shown the similarity in transcription machinery between archaea and eukaryotes in several aspects, such as subunit components, sequences of RNA polymerase, promoter

structure, and general transcriptional factors. These similarities led to the idea that the transcriptional regulation in archaea is closely related to that of eukaryotes. Nevertheless, the complete genome sequences of several hyperthermophilic archaeal species do not reveal homologs of the eukaryotic transcriptional factors, such as TBP-associated factors (TAFs), mediators, or coactivators/corepressors. Instead, the open reading frames of the putative bacterial transcriptional regulators were found in the genomes of these hyperthermophilic archaea (Bult et al. 1996, Klenk et al. 1997, Deckert et al. 1998). These findings indicate the possible existence of a mechanism of bacterial transcriptional regulation in archaea. Consistent with this, the recent study on a putative bacterial-like transcriptional regulator from *Sulfobobus solfataricus* has shown that this protein binds to the promoter of its own gene (Napoli et al. 1999). Another study on the putative transcriptional repressor from *Archaeoglobus fulgidus*, MDR1, demonstrated that MDR1 protein binds to operator sequences that overlap with its transcriptional initial site and represses its own gene expression. Moreover, this protein was shown to negatively regulate its own gene expression not by blocking the binding of TBP or TFB complex to its promoter, instead, by preventing RNA polymerase recruitment (Bell et al. 1999).

In the research presented here, the putative *lysR*-type transcriptional regulator gene from *M. jannaschii* has been subcloned and overexpressed in the *E. coli* cells. The purified recombinant protein is shown to bind specifically and selectively to the intergenic putative control region between the MJ *lysR*-type gene and a nearby upstream gene, although the function of this gene remains unknown. The precise binding position

of this protein was investigated by DNA footprinting analysis and the putative promoter region was estimated.

Materials and Methods

1. Plasmids, bacterial strains, and growth conditions

All plasmids and bacterial strains used in this study are described in Table 2-1. A map of the plasmid pET-5b is shown in Figure 2-1. All bacterial strains were grown in Luria-Bertani (LB) liquid medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, H_2O, 1 L) at 37°C with gentle shaking or on LB plates (LB with 15 g/L agar added). When necessary, ampicillin was added to a final concentration of 100 μg/ml. All water used was purified to >10 MΩ cm using a Milli Q water purification system (Millipore).

2. Primers and enzymes

Primers used in this study are depicted in Table 2-2. The relative positions of these primers are shown in Figure 2-2. All primers used in this study were synthesized by GIBCO BRL Custom Primer (Gibco BRL Life Technologies) and purified to standard level. The *Taq* DNA polymerase and all restriction enzymes were purchased from Promega. T4 DNA ligase and *pfu* DNA polymease were purchased from Stratagene.

3. Plasmid isolation and purification

The *E. coli* cells containing the desired plasmid were cultured in LB liquid medium containing 100 μg/ml ampicillin at 37°C with gentle shaking overnight. The cultured

cells were harvested by centrifugation (Sorvall RC-5B, GSA head, 4°C, 5000 rpm, 15 min). Wizard™ Mini/Midi/Maxipreps DNA purification kits were used for small-, medium-, and large-scale plasmid preparations (Promega). For a small-scale plasmid preparation, 3-4 ml of overnight culture of *E. coli* and a bench-top centrifuge (Model 5414, Brinkmann) were used. For a medium- or large-scale preparation, 40-60 ml or 200-300 ml of overnight cultures and a Sorvall RC-5B centrifuge (DuPont) were used. The detailed procedures are described in Promega Technical Bulletin (Promega). In a small-scale preparation, the disposable 3 ml syringes were used instead of a vacuum manifold. All purified plasmids in this study were stored at -20°C freezer.

4. Using polymerase chain reaction (PCR) to amplify the desired DNA fragments

4.1. Amplification the *lysR*-type gene from pAMBIJ46

The ORF of the *lysR*-type gene of *M. jannaschii* was amplified by PCR from a plasmid pAMBIJ46 using primer JK87 and JK88 (Table 2-2) and *pfu* DNA polymerase (Stratagene). The PCR was optimized by the modified Taguchi methods (Cobb & Clarkson 1994). The reaction was carried out in a 50 µl volume containing 4 U of *pfu* polymerase, 1x *pfu* polymerase buffer, 15 pmol of each primer, 30 ng template pAMBIJ46, 2.5 mM $MgCl_2$, 0.2 mM of 4 dNTPs, and water. The PCR thermal cycling conditions were 30 cycles of 94°C for 45s, 55°C for 1 min, 72°C for 1 min (Coy TempCycler/Model 50/60, Coy Laboratory Products Inc).

4.2. Amplification of the intergentic control region between the *lysR*-type gene and its upstream gene

The intergentic DNA fragment between the *lysR*-type gene and its upstream gene contains the putative control region for transcribing the both genes. The different segments of the DNA fragment in this region used in gel mobility shift assays and Dnase I footprinting analysis were amplified by PCR in a 50 μl volume, which consisted of 5 U *Taq* polymerase, 1x *Taq* polymerase buffer, 200 μM 4 dNTPs, 20 ng pAMBIJ46, 1 μM each primer, 2.5 mM $MgCl_2$, and water. The thermal cycling reaction conditions were 30 cycles of 94°C for 40 s, 55°C for 40 s, 72°C for 40 s on Coy TempCycler (Model 50/60).

5. Purification of PCR products

The PCR product was first treated three times with StrataClean™ resin to remove DNA polymerase (Stratagene). To do so, seven μl of resin were combined with 50 μl of PCR reaction, followed by vortexing the mixture for one minute. The mixture was incubated at room temperature for one minute and the resin was removed by centrifugation. Then a Wizard™ PCR Preps DNA purification kit (Promega) was used to further purify the PCR products. The direct purification from PCR reactions was carried out by using disposable 3 ml Luer-Lok® syringes in this study. The purified PCR products were stored at -20°C freezer.

6. Modified Taguchi method to optimize PCR reaction

The modified Taguchi method was used to optimize the polymerase chain reaction (Cobb & Clarkson, 1994). Although the basic method for PCR is well established, a novel application might require an investigation of a range of reaction components, which could alter yields, specificity, or reaction fidelity. Taguchi methods are widely used in industrial process design. The modified Taguchi method was demonstrated to be suitable for optimizing PCR by revealing the effects and interactions of specific reaction components simultaneously. In our experiment, 9 PCR reactions were carried out at three levels with primer at 7.5/15/30 pmoles, the template at 10/20/30 ng, $MgCl_2$ at 0.5/1.0/5.0 mM, and 4 dNTPs at 0.05/0.2/0.4 mM. The product yields (Table 2-4) for each reaction were used to estimate the effect of individual components on amplification by using the quadratic loss function.

$$SNL = -10\log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right]$$

where SNL is the signal-to-noise ration, n is the number of levels, and y is the product yield. For each reaction component, the optimal conditions are those that give the largest *SNL* (Table 2-5). The reaction could be further refined by plotting a polynomial regression curve from *SNL* values versus the amount of each component and by choosing the level that maximizes *SNL* (Cobb & Clarkson, 1994).

7. Double restriction enzyme digestion

The purified PCR product and plasmid vector pET-5b were both subjected to restriction enzyme digestions to produce sticky ends. The 50 µl reactions were carried out in 1.5 ml Eppendorf tubes containing 100-200 ng DNA, 1x reaction buffer, 2-3 U of the first restriction enzyme (*Bam*HI). The reactions were mixed well and incubated at a 37°C for 15-20 min. The digestion was checked with 1% agarose gel electrophoresis. If the digestion was complete, 2-3 U of the second restriction enzyme (*Eco*RI) were added, then mixed thoroughly and incubated at a 37°C for another 15-20 min. The reactions were stopped by adding 0.5 M EDTA (pH 8.0) to a final concentration of 10 mM. If the digestion with the first enzyme was incomplete, more of the first enzyme was added to continue incubation until the digestion was complete before the second restriction enzyme was added.

8. Purification of DNA fragments digested by restriction enzymes

The DNA fragment digested by restriction enzymes (*Eco*RI and *Bam*HI) has to be purified before being subjected to a ligation reaction. The fragment was first heated at 70°C for 15 min to deactivate the restriction enzymes, then StrataClean™ resin (Stratagene) was used to remove them. Finally, the DNA solution collected from StrataClean™ resin purification was filtered through a Millipore ultrafree-MC filter to remove salts and other small molecules. The purified DNA was then stored at -20°C.

9. Ligation of the *lysR*-type gene from *M. jannaschii* to the plasmid vector pET-5b

9.1. Checking availabilities of sticky ends of the vector pET-5b and the insert *lysR*-type gene

The availabilities of sticky ends of both the vector and the insert were checked by a self-ligation reaction. The self-ligation reaction was carried out in a 15 µl volume with 30-50 ng DNA, 1x ligase buffer, 1 U of T_4 DNA ligase, and water. The mixtures were incubated at room temperature overnight. The reaction was resolved on a 1% agarose gel and observed under UV-light.

9.2. Ligation of vector and insert DNA

One hundred ng of the double-cut vector *p*ET-5b were used in the ligation reaction. The amount of the insert DNA used was calculated with the following formula:

$$ng\ of\ insert = \frac{ng\ of\ vector\ x\ kb\ size\ of\ insert}{kb\ size\ of\ vector}\ x\ molar\ ratio\ of\ insert/vector$$

A 1:3 molar ratio of vector (4.1 kb) and insert (0.9 kb) was used. About 60 ng of the insert were needed in the ligation reaction. The reaction was carried out as described in Promega pET-5b expression vectors technical manual (Promega).

10. Preparation of *E. coli* competent cells

The JM109 and BL21(DE3)pLysS competent cells for transformation were prepared according to the modified $CaCl_2$ method (Tang et al. 1994). The other method used to prepare the competent *E. coli* cells is the modified TSS protocol

(http://epicentre.com/f2_3/f2_3ts.htm). The original TSS method was described in Chung's paper (Chung et al., 1989).

11. Transformation of competent *E. coli* cells

11.1. Determination of transformation efficiency of the competent cells

Ten ng of supercoiled pET-5b vectors were used to transform 100 µl of the competent cells prepared above. Aliquotes of 1, 10, and 25 µl of the transformation culture were plated on LB plates containing 100 µg/ml ampicillin and incubated at 37°C overnight. The transformation efficiency, expressed as the number of transformants per microgram of DNA, is calculated with the following formula:

$$Efficiency = \#\ of\ transformed\ colonies\ x\ \frac{10^5}{volume\ of\ aliquots(ul)}$$

The transformation efficiencies of the competent cells prepared by the above methods were in the range of 10^5-10^7.

11.2. Transformation of the competent cloning host JM109 cells

The procedure was described in pET-5 expression vectors technical manual (Promega) with the exception of using a 100 µl aliquot of the competent JM109 cells. The transformed cells were plated on LB plates containing 100 µg/ml of ampicillin and grown at 37°C. The colonies growing on the plates were collected and inoculated into 5 ml LB/ampicillin liquid medium. The overnight cultures were centrifuged (Model 5414,

4°C, 5000 rpm, 10 min) and resuspended into 5 ml of LB-ampicillin liquid medium with 25% glycerol, and stored at –70°C.

12. Isolation, purification and verification of recombinant plasmid DNA

The cloning host JM109 *E. coli* cells carrying the recombinant plasmid pET-5b were grown in LB liquid medium with 100 μg/ml ampicillin at 37°C overnight. The recombinant pET-5b carrying the insert, *lysR*-type gene from *M. jannaschii,* was isolated and purified as described above. The purified plasmids were digested by *Eco*RI and *Bam*HI to verify the existence of the insert *lysR*-type gene. To screen for the wild-type *lysR*-type gene, an automated DNA sequencing was carried out. To do so, primer JK89 and JK88 were used to sequence the insert *lysR*-type gene from both ends using the recombinant pET-5b as a template. The sequencing reaction consists of 400 ng of the purified recombinant plasmid, 3 μl of Prism™ ready sequencing mix, 3 pmol of each primer, and water to arrive a final volume of 15 μl. Cycle sequencing was carried out in a 0.2 ml thin wall tube on a Perkin-Elmer GeneAmp® PCR System 9600 thermal cycler. The reaction conditions were as follows: 1 cycle of 96°C/1 min; 25 cycles of 96°C/10 s; 50°C/5 s; 60°C/4 min. Excess nucleotides were removed by passing the products though Sephadex G-50 (Sigma) spin columns (Princeton Separations Centri-Sep™). Samples were resolved on a 4.75% polyacrylamide gel by an ABI 373 DNA Sequencer. Sequence data were then analyzed with DNA Sequence Navigator and GenePro, 6.10.

13. Transformation of expression host BL21(DE3)pLysS *E. coli* competent cells by the recombinant pET-5b carrying the wild-type *lysR*-type gene

The purified recombinant pET-5b carrying the wild-type *lysR*-type gene was used to transform the competent expression host BL21(DE3)pLysS *E. coli* cells as described in 11.2. above.

14. Induction of gene expression by Isopropyl-β-D-thiolgalactoside (IPTG)

The BL2(DE3)pLysS cells with the recombinant plasmid pET-5b carrying the wild-type insert were grown at 37°C in LB liquid medium containing 100 μg/ml ampicillin. The next day 1-2 ml of the overnight culture were removed and inoculated into 100 ml fresh LB medium containing 100 μg/ml of ampicillin. The cultures were grown at 37°C until OD_{600} is about 0.5-0.6. The 0.1 M IPTG stock solution was added to the log phase culture to arrive a final concentration of 0.4 mM. The cultures were continued to incubate at 37°C for an additional 3-4 hours. Overnight incubation may increase the yields of the induced protein.

15. Protein isolation and purification

The BL2(DE3)pLysS cells in which the expression of the *lysR* gene was induced were pelleted by centrifugation (5000 rpm, 20 min, GSA head, Sorvall RC5B) at 4°C. The pellets were resuspended in the sonicated buffer (20 mM tris-HCl at pH 7.5, 1 mM EDTA, 200 mM NaCl). Sonication was carried out with a Sonication Ultrasonic Processor (Heat Ultrasonics System, Model 225) on ice for 8 cycles, 30 s each cycle at

1 min intervals (40% duty cycle at a power setting of 4). The cell rupture was considered complete when the lysate became clear or the protein concentration measured by Bradford reagent assay (Bradford 1976) showed no significant change. The cell debris was removed by centrifugation (14000 rpm, 20 min, SS-34 head, Sorvall RC5B) at 4°C. The clear extract was transferred to 50 ml sterile centrifuge tubes and pellets were discarded. The protein concentration of the crude solution was measured by Bradford reagent assay. A 3 ml aliquot of the clear extract was added into a 5 ml sterile Eppendorf tube and heated at 85°C for 20 min. The heat unstable proteins were spun down at the top speed on a bench-top centrifuge (model 5414, Brinkmann) at room temperature. The supernatant was decanted to a 15 ml sterile centrifuge tube and stored at 4°C. The protein concentration at this stage was measured by Bradford reagent assay.

Bio-gel 0.5 m (Bio-Rad) size exclusion matrix was packed into a glass Bio-Rad chromatography column (36.5 cm x 1.5 cm). The column was pre-equilibrated with the column buffer (50 mM tris-HCl at pH 7.5, 2 mM EDTA, 100 mM NaCl). The 1 ml supernatant treated by heating was loaded onto the pre-equilibrated column with a 4-way sample injector. The column was eluted with the column buffer with a pump setting at 65. The fractions were collected with a time setting at 3.0 min. The protein concentration of each fraction was determined by Bradford reagent assay. The fractions were analyzed on a SDS-PAGE gel (Bio-Rad Mini-PRPTEAN® II Dual Slab Cell). The SDS-PAGE gel was scanned by IS-1000 Digital Image System (AlphaImager™ Version 3.2, Alpha Innotech Corporation) to locate the expressed LysR-type protein band and determine the purity of it. In addition, gel mobility shift assays were carried

out to check the binding ability of the purified protein. The fractions with the highest purity and binding activity were pooled and concentrated by Ultrafree®-MC Centrifugal Filter Units (NMWL 30,000, Sigma). All protein solutions were stored in 4°C refrigerator.

The molecular weight (MW) of native MJ-LysR protein was estimated by gel filtration with the same column that was calibrated with thyroglobin (MW 670,000), gamma globulin (MW 158,000), ovalbumin (MW 45,000), myoglobin (MW 16,950), and vitamin B-12 (MW 1350) obtained from Sigma.

16. Specific binding activity of purified LysR-type protein

To test the specific binding activity of the purified MJ-LysR protein, a gel mobility shift assay was carried out with several controls. A 555 bp of the intact intergenic DNA fragment, which contains the first 131-bp of the *lysR*-type gene, the entire putative control region between the *lysR*-type gene and its upstream gene, and the first 61 bp of the upstream gene, was amplified by PCR using primer 90 and primer 91 as described in 4.2. above. The PCR products were purified as described in 5. The conditions for the binding reaction were as follows. A 10 μl reaction volume contains 30 ng of the purified intergenic DNA fragments from *M. jannaschii* or the *lysR*-type intergenic DNA fragments from *Burkholderia cepacia*, 70 ng of the purified MJ-LysR protein or bovine serum albumin (BSA) serving as a control protein, 2 μl of 5X binding buffer (100 mM Tris at pH 7.5, 25 mM NaOAc, 25% glycerol, 0.005% BSA), and water. The reactions were incubated at room temperature for 5 min to allow the proteins to interact with the

DNA. Following the protein-DNA binding, 1 μl of dye solution (0.25% bromophenol blue, 40% sucrose) was added to the mixtures, the samples were loaded onto a 4.5% polyacrylamide non-denaturing gel (29.2:0.8 acrylamide to bisacrylamide in 50 mM tris, 380 mM glycine, 2 mM EDTA, and 2.5% glycerol, pH 8.0). Electrophoresis was carried out at 4°C at 97 V in 1X TBE running buffer (45 mM Tris-borate, 1 mM EDTA). After electrophoresis, gels were stained in a staining solution (100 ml of 1X TAE buffer, 20 μl of 0.55 mg/ml ethidium bromide) for 30 min and rinsed with water several times. The gels then were scanned and analyzed by using IS-1000 Digital Image System (AlphaImager™ Version 3.2, Alpha Innotech Corporation).

17. Binding activity of the purified MJ-LysR protein at different temperatures

To investigate the binding activity of the purified MJ-LysR protein at different temperatures, a gel mobility shift assay was carried out in two ways. First, the protein-DNA mixtures were prepared as described in 16 and then incubated at different temperatures: 0°C, 25°C, 37°C, 65°C, 85°C, and 95°C for 10 min. Following the incubation, electrophoresis and visualization of gel were conducted the same as in 16.

The thermal stability of the MJ-LysR protein was measured by incubating the protein solutions at 0°C, 25°C, 65°C, 79°C, and 95°C for 30 min first. After the heated protein solutions were quenched on ice, 70 ng of the heat-treated protein were incubated with 30 ng the intergenic DNA fragment from *M. jannaschii* in 1x binding assay buffer (20 mM tris at pH 7.5, 5 mM NaOAc, 5% glycerol, 0.001% BSA) at room temperature for 5 min. The mixtures were resolved on a 4.5% polyacrylamide non-denaturing gel

(29.2: 0.8 acrylamide to bisacrylamide) as described in 16. Electrophoresis and visualization of the gel followed steps in 16.

18. Estimation of approximate binding region on the intergenic DNA fragment

The 555 bp intergenic DNA fragment was amplified by PCR using primer JK90 and JK91. The PCR reaction and product purification were addressed in section 4.2. and section 5 in this chapter.

The approximate binding region of the MJ-LysR protein on the intergenic DNA fragment was estimated by the restriction enzyme digestions combined with gel mobility shift assays. The cutting sites of the restriction enzymes on the DNA fragment are shown in Figure 2-3. The restriction enzymes, *Ssp*I, *Alu*I, *Rsa*I, and *Bam*HI, were purchased from Promega Corporation. Double and single restriction enzyme digestion reactions were followed by removing the restriction enzymes from the reaction mixtures with StrataClean™ resin and by removing salts with Millipore ultrafree-MC filters. The purified DNA fragments pre-treated by the restriction enzyme(s) were incubated with the purified MJ-LysR protein as described in section 16, and followed by a 4.5% polyacrylamide non-denaturing gel electrophoresis and analysis by IS-1000 Digital Imaging System.

According above results, several pairs of primers were designed to amplify DNA fragments containing different segments of the intergenic region. Gel mobility shift assays were conducted with these fragments incubating with the MJ-LysR protein as

described in section 16 to locate a more precise binding region on the double-stranded DNA.

19. Dnase I footprinting analysis

From the previous results, the putative binding site of the MJ-LysR protein can be estimated. Two pairs of primers were designed and used to synthesize a 229-bp fragment from the coding strand (JK97/JK91) and a 330-bp fragment from the noncoding strand (JK94/JK93). The purified PCR fragments were used in Dnase I footprinting analysis.

19.1. Preparation of radioactive labeled DNA fragments for DNA footprint analysis

Primer JK97 and JK93 were labeled using [γ-^{32}P] ATP (6,000 Ci/mmol) (Amersham Pharmacia Biotech) and T4 Polynucleotide Kinase (Promega Corporation). The labeling reaction containing 50 pmoles of oligonucleotides, 50 pmoles of [γ-^{32}P] ATP (6,000 Ci/mmol), 3 U of the kinase, and water, was carried out in a 50 µl volume. The mixtures were incubated at 37°C for 15 min. The reaction was quenched by adding 2 µl of 0.5 M EDTA. The kinase was removed by StrataClean™ resin (Stratagene).

The DNA segments containing the putative binding site for the MJ-LysR protein were amplified by PCR as described above except using one "hot" (radioactive labeled) primer and one "cold" (non-radioactive labeled) primer. The PCR product was cut from a 1% agarose gel and purified by using Quantum Prep® Gel Slice Kit (Bio-Rad).

19.2. Gel preparation

Sequi-Gen® Nucleic Acid Sequencing Cell was used for electrophoresis (Bio-Rad Laboratories). Before assembling the glass plate sandwich, both plates were thoroughly cleaned by scrubbing with powdered detergent (Alconox), washed thoroughly with tap water, and finally, rinsed with deionized water. The plates were dried with Kimwipes® wipers. The back plate needed to be coated occasionally (once a month if used 2-3 times/week) with Sigmacote® (Sigma) to prevent gel sticking. Several milliliters of Sigmacote® were spread over the plate evenly, dried in fume hood for 30 min, and the excess rinsed thoroughly with deionized water. The front plate was treated with siloxy coating (0.8 ml 95% ethanol, 4.2 ml 3-(trimethoxysilyl) propyl methacrylate (Aldrich chemical company, Inc), and 25 µl of 10% HOAc) to help sticking. The siloxy coating solution was spread evenly over the plate and the plate was baked for 5 min in an oven at medium setting. After the plate was cooled to room temperature, rinsed with lots of water, and dried with Kimwipes. The gel apparatus was set up as described in Sequi-Gen® Nucleic Acid Sequencing Cell instruction manual (Bio-Rad Laboratories).

One hundred twenty five milliliter of 10% sequencing gel solution were prepared (52.5 g urea, 12.5 ml 10x TBE, 41.7 ml 30% acrylamide stock solution and H_2O to 125ml), filtered through 0.45 µm filter paper, and degased. Thirty five milliliter of the gel solution made by mixing 196 µl of 25% ammonium perphosphate (APS) (Sigma) with 145 µl of TEMED (Sigma) were taken to seal the bottom edge of the gel. To the rest of the gel solution, 108 µl of APS and 72 µl of TEMED were added. The gel

solution was injected into the plate sandwich slowly to avoid air bubbles. A comb was placed between the plates before the gel polymerized.

19.3. DNA footprint reactions

The incubation reaction was performed in a 0.5 ml sterile Eppendorf tube containing about $5x10^4$ cpm singly labeled DNA fragment, amount of the MJ-LysR protein varying from 0 ng to 75 ng, 3 µl of 5x assay buffer (100 mM tris at pH 7.5, 50 mM NaOAc, 2 mM $CaCl_2$, 0.05% BSA, 5% glycerol), 500 ng of herring sperm DNA, and H_2O to a final volume of 13 µl. The mixtures were incubated at room temperature for 15 min. Approximately 0.25 U of RQ1 Rnase-Free Dnase I (Promega Corporation) was added to each reaction, the mixture was incubated at 37°C for 40 sec, and then the reaction was stopped by adding 2 µl of 0.5 M EDTA. The mixtures were kept on ice until all reactions were completed. The 300 µl of H_2O was added to each tube and all reactions were purified with Ultrafree-MC filter (NMWL 10,000, MilliPore). Again all the reactions were precipitated in 100% ethanol and washed twice with 70% ethanol. The precipitated DNA was dried with a speed vacuum (Labconco) and resuspended in 4 µl loading buffer (New England BioLabs Inc.). The samples were loaded on a 10% pre-electrophoresed DNA sequencing gel.

To obtain sequencing information on the intergenic fragment, DNA sequencing was carried out on the intergenic fragment by using chemical methods (Ausubel et al. 1995).

19.4. Electrophoresis

Electrophoresis was carried out with a Bio-Rad model 3000/300 power supply. The gel was pre-electrophoresed at a constant power 74 W (about 2000 V) for about one hour. The sample electrophoresis was performed at the same voltage for about 3-4 hours until the dye band reached the bottom of the gel. The gel was fixed in a fixer solution (10% methanol and 10% acetic acid) (Sigma) for 30 min and dried in an oven at a medium setting for 5 min. In a darkroom, an autoradiograph was established by exposing the gel to X-ray film (XAR-351, Kodak Scientific Imaging Film) for 24 hours at room temperature. The film was developed and fixed with Kodar developer and fixer for 4 min each, and rinsed between each operation and after fixation.

20. Prediction of promoter elements in the intergenic DNA region

To search for the potential binding sites of the general transcription factors and the promoter element on the intergenic fragment, the conserved consensus TATA box sequence in archaea, 5' $^{T}_{A}$ $^{T}_{A}$TATATA-3' (Palmer & Daniels 1995, Gelfand et al. 2000) preceded by a run of 3 or 4 purines, was used as a probe. The matches on the intergenic putative control region between MJ *lysR*-type gene and its upstream gene were considered as the putative promoter region. This was done visually and by the Omiga multi-alignment program.

Results

1. Cloning of *lysR*-type gene

The *lysR*-type gene was amplified from pAMBIJ46 by PCR. Our PCR reaction was optimized and concluded with 5 pmol of each primer, 30 ng template, 5 mM $MgCl_2$, 0.2 mM of 4 dNTPs (Table 2-5).

The PCR product, the MJ *lysR* gene, was ligated into the plasmid vector pET-5b at the *Bam*HI and *Eco*RI sites. The recombinant plasmids were used to transform the competent cloning host *E. coli* JM109 cells. The transformants were collected for plasmid isolation. The isolated and purified plasmid was digested by the *Bam*HI and *Eco*RI to check for the existence of the insert *lysR*-type gene. The results were showed in Figure 2-4. About half of the recombinant vectors contained the insert *lysR*-type gene. An automated DNA sequencing was performed to verify the identity of the wild-type insert gene. Plasmid carrying the wild-type *lysR*-type gene was successfully transformed the competent expression host *E. coli* BL21(DE3)pLysS cells. The transformants were collected for protein induction and storage.

2. Protein expression and purification

The *lysR*-type gene was inserted downstream from the IPTG inducible *lacUV5* promoter of plasmid vector pET-5b. The gene expression was successfully induced by addition of 0.4 mM IPTG. The full length of MJ-LysR fusion protein, consisting of 296 amino acids and 11 extra amino acids at the N-terminus derived from the vector pET-5b, was purified through two steps: heat treatment at 85°C for 20 min, followed by

size exclusion chromatography. The heat treatment removed most of the heat sensitive proteins. The protein purity at this stage is about 70% homogenous judging from SDS-PAGE analysis (Figure 2-5). Size exclusion chromatography improved the protein purity to nearly 85-90% (Figure 2-5).

The molecular mass of native MJ-LysR protein was estimated to be 75 KD using five protein standards as references in gel filtration, which is in close agreement with the mass (36.8 KD for monomer) predicted from the DNA sequence of the fusion *lysR*-type gene, indicating that MJ-LysR protein exists as a dimer in solution.

3. Binding of MJ-LysR protein to its upstream DNA fragment

A gel mobility shift assay was used to analyze the specific binding activity of the MJ-LysR protein. In this experiment, BSA was used as a protein control and two intergenic DNA fragments with different lengths from *Burkholderial cepacia* as DNA controls. Both DNA fragments from *Burkholderia cepacia* contain the promoter region of the *dgdA* gene that is under the control of another LysR-type transcriptional regulator, DgdR. It was concluded that the purified MJ-LysR protein bound to its upstream intergenic putative control region specifically and selectively based on the following evidence (see Figure 2-6): (1) BSA did not bind to the intergenic DNA containing the putative control region for the *lysR*-type gene and its upstream gene from *Methanococcus jannaschii* as shown in lane 2. (2) The purified MJ-LysR protein did not bind to the promoter region of the *dgdA* gene as shown in lane 3 and lane 4. (3) The

purified MJ-LysR protein did bind to the intergenic control region amplified from *M. jannaschii* as shown in lane 6 and lane 7.

4. Thermal stability of MJ-LysR protein

To check the thermal stability of the purified MJ-LysR protein, two experiments were carried out. In the first, the protein was mixed with the DNA fragment before incubating at different temperatures. In the second experiment, the purified MJ-LysR protein was pre-treated at 0°C, 25°C, 37°C, 64°C, 79°C, and 94°C for 30 min, followed by incubating with the DNA fragments at room temperature. The DNA- protein complex was resolved on a 4.5% polyacrylamide non-denaturing gel. The results were shown in Figure 2-7 and Figure 2-8, respectively. Figure 2-7 shows that the more of the free DNA was converted to the shifted bands with the increasing of the temperature until 85°C. In other words, the binding activity of the MJ-LysR protein increased with temperature rising at least to 65°C. At 85°C, only one shifted band can be observed. At 94°C, no shifted band could be observed at all on the gel. These two results suggested that the binding activity of the MJ-LysR protein decreased but still was partially maintained at 85°C, but totally lost at 94°C. In contrast, with increasing temperature, the proportion of the shifted bands was increased until 79°C (Figure 2-8). Even after the MJ-LysR protein was heated at 94°C for 30 min, one shifted band could be observed clearly on the gel in Lane 6. The results indicate that the binding activity of the MJ-LysR protein increased with increasing temperatures until 79°C and at least 75% of its binding activity was still maintained at 94°C.

5. Estimation of approximate binding region of MJ-LysR protein

The approximate binding region of the MJ-LysR protein was estimated by combining restriction enzyme digestions with a gel mobility shift assay. The 555 bp DNA fragment, including the entire intergenic region between the *lysR* gene and its upstream gene, partial of the *lysR* gene, and its upstream gene, was first digested by restriction enzymes. The purified DNA mixture then was incubated with the MJ-LysR protein at room temperature. The results are shown in Figure 2-9, Figure 2-10, and Figure 2-11. The lane 2 in Figure 2-9 shows the incomplete digestion of *Alu*I. The fast-moving band should contain two DNA fragments, sized 277 bp and 278 bp, separately. The amount of the total DNA in these two bands is about 8.9 ng calculated by using AlphaImage 3.2 System (See Appendix I and II for detailed calculation). In the lane 3, the purified mixture was incubated with the purified MJ-LysR protein. As shown in the lane 3, one or both DNA fragments present in the fast-moving bands made band shifting. The amount of free DNA fragment now reduced to 6.6 ng calculated by using the same software and method. The uncut DNA fragment (the slow-migration band in lane 2) also made band shifting due to the presence of the binding site of the MJ-LysR protein on it. In the lane 4, the complete digestion of the 555 bp intergenic DNA fragment by *Alu*I and *Bam*HI generated three bands, 130 bp, 147 bp and 278 bp while the incomplete digestion of the *Bam*HI fragment by *Alu*I gave a fragment sized by 425 bp. When the purified mixture was incubated with the purified MJ-LysR protein, two more shifted bands were generated. The fast-moving new band is thought from the binding of the MJ-LysR protein to the 147 bp DNA fragment that is the second band

from the bottom of the gel. The amount of the 147 bp DNA fragment decreased from 3.5 ng to 1.7 ng. The amount of DNA in 130 bp and 278 bp bands remain unchanged. The slow-moving new shifted band is from the incomplete digested 425-bp DNA fragment since the amount of the 425 bp DNA fragment decreased significantly, too (from 12.2 ng to 4.7 ng). Figure 2-10 shows the results of *Rsa*I/*Bam*HI digestion and gel mobility shift assay of the mixtures. Two fragments, 338 bp and 217 bp, were generated from *Rsa*I digestion, while combining with the purified MJ-LysR protein, only the 217 bp fragment produced two shifted bands. The band intensity of the 338 bp fragment did not have a significant change while the amounts of the 217 bp fragment reduced from 13.7 ng to 4.7 ng. In the Figure 2-10, *Bam*HI digestion of the intergenic DNA fragment generated two bands: 130 bp and 425 bp (lane 4). As shown in lane 3 of Figure 2-10, when the band mixtures were combined with the MJ-LysR protein, only the 425 bp fragment made gel shifting. The quantity of the 425 bp band changed from 22.5 ng to 11.7 ng. Figure 2-11 showed the results from *Rsa*I and *Ssp*I digestion. The complete digestion of *Rsa*I and *Ssp*I generated three bands, 102 bp, 217 bp, and 236 bp showing in lane 2. The newly generated shifted band in lane 3 is thought from the binding of the MJ-LysR protein to the 217 bp fragment since the amount of which changed significantly, from 10.7 ng to 5.0 ng. The *Ras*I digestion generated two bands of size 217 bp and 338 bp. The newly formed two shifted bands in lane 5 were from the 217 bp fragment due to the significant changes in amount of the DNA (from 10.3 ng to 5.5 ng). A summary of these results is shown in Figure 2-12. From Figure 2-12, it was

concluded that the region between the cutting sites of restriction enzyme *Rsa*I and *Bam*HI should be where the MJ-LysR protein bound.

To confirm the above results, primer JK94/JK93, JK92/JK91, JK90/JK96, and JK90/JK91 were used to amplify the different segments of the intergenic DNA fragment from pAMBIJ46. The purified DNA segments were used to carry out gel mobility shift assays. The results showed that the segments amplified by primers JK92/JK91, JK94/JK93, and JK90/JK93 contain the MJ-LysR protein binding site while the segment amplified by JK90/JK96 does not. It is interesting to note that the two shifted bands generated by the binding of the MJ-LysR protein to the 307 bp segment, which was amplified from primer JK92/JK91, are located at about the same positions as those generated by the binding of the MJ-LysR protein to the 320 bp segment amplified from primer JK94/JK93 as shown in Figure 2-13. Normally, the shifting bands generated from the binding of the protein to the shorter fragment should be moving faster through the gel than those generated from the binding of the same amount of the protein to the longer fragment if there is no bending involved. The possible reason for the retardation of the shifted bands generated from the binding of MJ-LysR protein to the 307 bp segment is that the binding site on this segment might be closer to the center compared to what it is on the 320 bp segment. With the binding site being closer to the center, the distance of the both ends is shorter and leads to slower migration through the gel than having the binding site at either end of the fragment with approximately same length. This result provides evidence of the existence of bending. A summary of these results indicates that the region between primer JK96 and JK93 is crucial for the binding of the

MJ-LysR protein (Figure 2-15). These results are coincident with the restriction enzyme tests and further confirmed the conclusion derived from the restriction enzyme digestion experiments, at the same time, they provide solid evidence on the bending induced by the MJ-LysR protein.

6. DNase I footprint analysis

DNase I footprint analyses were performed to locate the precise binding site of the MJ-LysR protein on the intergenic region. A 229 bp DNA fragment from the coding strand and a 330 bp DNA fragments from the noncoding strand were used. Comparisons of the sequence patterns produced in the absence and in the presence of the MJ-LysR protein demonstrated a protected region of 30 nucleotides on the coding strand and 29 nucleotides on the noncoding strand, respectively (Figure 2-16). The binding site extended from position –20 to –50 on the coding strand and from position –25 to –54 on the noncoding strand relative to the translation start point of the *lysR*-type gene (Figure 2-16). With each strand, one weak Dnase I hypersensitive site was present: at position –31 on the coding strand and at position –43 on the noncoding strand. A summary of Dnase I footprint data is shown in Figure 2-17. The protected region on each strand was indicated by a bracket and the Dnase I hypersensitive sites were labeled by asterisks.

7. Prediction of promoter elements in the intergenic region

To search the potential promoter elements, a conserved consensus TATA box sequence in archaea, 5' $^{T}_{A}$ $^{T}_{A}$TATATA-3' preceded by a run of at least 3 purines was used

as a probe to search for the matches on the intergenic region. Several matches were found on the coding strand. One of them is located at –51 and extended to –41 on coding strand relative to the translation start site of the MJ-LysR protein. This is the only one that is preceded by a run of three purines. This region is thought to be the core promoter region that controls the MJ *lysR* gene expression. Only match found on the noncoding strand is located at –285 and extended to –295 indicated by capital letters preceded by three adenines (Figure 2-18). All the matches on the coding strand were underlined. According to the position of the box A and the position of box B relative to the open reading frame in other archaeal species, the match located at base pair position from –51 to –41 is most likely to be the promoter region for controlling the expression of the MJ *lysR*-type gene. The match located at base pair position from –285 to –295 is the potential promoter region for the gene with unknown function located upstream from the MJ *lysR* gene in this study.

Discussions

1. Cloning and expression of the *lysR*-type gene from *M. jannaschii*

To successfully clone the *lysR*-type gene from *M. jannaschii*, it is very important to generate high concentration of the sticky ends. These sticky ends were produced by *Bam*HI and *Eco*RI digestions. Instead of doing two enzyme digestions at the same time, the insert fragment and the vector were digested by one enzyme first, usually the one with the lower cutting efficiency. After the digestion by the first enzyme was ensured to be complete, the second enzyme was added into the reaction. This would increase the

efficiency of ligation in the following steps by increasing concentrations of the sticky ends in both the insert fragment and the vector, at the same time decreasing the ratio of self-ligation by decreasing amount of single-cut vector. It is especially important to do the digestion separately when the cutting efficiency of one of the enzymes is lower. A self-ligation test was designed to check the availabilities of the sticky ends in the insert fragment and the vector. The availabilities of the sticky ends were viewed on 1% agarose gel with a pattern of ladder after it was ligated by ligase (Figure 2-19). The self-ligation test is also useful for the blunt end ligation. It is harder to ligate the blunt end fragment due to the nature of the blunt ends. The result of the ligation reaction can be predicted beforehand by checking the availabilities of the both ends with a self-ligation test.

The unsuccessful cloning of the MJ *lysR*-type gene into the original vector system, pBTac, led to a hypothesis that the product of this MJ *lysR*-type gene might have a toxic effect on its cloning host and kill the host cells. To avoid potentially toxic effects of the MJ *lysR*-type gene, the pET vector system was chosen for protein production. The pET-5b vector, derived from pBR332, contains the promoter and translation start site from T7 bacteriophage, polylinker site including *Bam*HI and *Eco*RI cloning sites, and the ampicillin resistance gene (Figure 2-1). The MJ *lysR*-type gene was initially cloned into the pET-5b in JM109, which does not carry the gene for T7 RNA polymerase. This has significant meaning for the successful cloning of a toxic gene. The products of toxic genes could kill the cloning strain during transformation, thus leading to a failure of cloning process. However, in the pET expression system, the cloned gene was

expressed under the transcriptional control of a bacteriophage T7 promoter. The transcription of the cloned *lysR*-type gene remained silent due to a lack of the T7 RNA polymerase gene in the cloning host JM109. Following verification of the construct by restriction enzyme analysis and DNA sequencing, the recombinant plasmid vector containing the wild-type MJ *lysR*-type gene was transformed into the expression host BL21(DE3)pLysS cells. This strain is a lysogen of bacteriophage DE3, a lambda derivative that contains the gene for T7 RNA polymerase under the control of the IPTG inducible *lacUV5* promoter. Nevertheless, in this expression strain, a low level of the expression of T7 RNA polymerase gene exists even in the absence of IPTG. A plasmid pLysS that expresses low levels of T7 lysozyme was present in the BL21(DE3)pLysS strain to prevent the potential basal transcription. The T7 lysozyme is a natural inhibitor of T7 RNA polymerase gene. The amount of T7 lysozyme produced by this plasmid is sufficient to inhibit the expression of T7 RNA polymerase, but is too low to prevent the gene expression in the presence of IPTG. Therefore, it is possible to clone and express the toxic genes in the pET expression system.

The *lysR*-type gene from *M. jannaschii* was subcloned into the plasmid vector pET-5b. The resulted recombinant gene contains 33 extra base pairs in front of the original *lysR*-type gene, which led to a fusion protein that has 11 extra amino acids at the N-terminus after translation (Figure 2-20). Do these extra amino acids affect the binding activity of MJ-LysR protein? As mentioned above, the N-terminal amino acids in the LysR family adopt a helix-turn-helix secondary structure that is considered the most used DNA recognition and binding motif in bacteria. The extra N-terminal amino

acids may interfere with the binding activity of the protein by introducing a long tail adjacent to the helix-turn-helix motif. However, the experiments in this study showed that this fusion protein does bind to the intergenic DNA containing the putative control regions for both genes.

A definitive answer to this question, does the extral amino acids at the N-terminal affect the binding of the MJ LysR protein is also possible by comparing the fusion protein with other LysR proteins. The multiple alignment of the N-terminal amino acid sequences of other members of the LysR family shows that some LysR-type proteins also possess long tails adjacent to the helix-turn-helix motif (Figure 2-20). Even with the 11 extra amino acids, the fusion MJ-LysR protein is still one amino acid shorter than the DsdC protein from *E. coli* and one amino acid longer than the DgdR protein from *B. cepacia*. It was concluded from these observations that 11 extra amino acids at the N-terminus of the MJ-LysR protein will not, or at least, will not completely prevent the binding activity of this protein.

2. Binding activity of MJ-LysR protein.

The results from the gel mobility shift assay demonstrated the specific and selective binding activity of the MJ-LysR protein to the intergenic DNA region, which contains the putative control regions for regulating the expression of both the MJ *lysR*-type gene and the adjacent upstream gene (Figure 2-6). Binding in this region indicated that the MJ-LysR protein has the potential to regulate the both gene expression divergently and simultaneously, which is consistent with the regulation mode of other LysR family

members (Schell 1993). It is interesting to note that the MJ-Lys protein did not recognize and bind to the intergenic control region of the *dgdA* and *dgdR* gene from *Burkholderia cepacia* (lanes 3 and 4 Figure 2-6), although both regions contain a consensus DNA sequence, $TN_{11}A$, that is present in recognition sites for other LysR-type regulators (Schell 1993). The specific binding of the MJ-LysR protein to its own intergenic region despite the similarity of the recognition site from other species might indicate the sequence-specific recognition and binding mode for members in the LysR family. Furthermore, upon the binding of the MJ-LysR protein, the two retarded bands were observed on a non-denaturing polyacrylamide gel (Figure 2-6). This result indicates the formation of DNA/protein complex. The faster migration band might be resulted by a complex of DNA and the MJ-LysR dimer while the slower migration band, a complex of DNA and the tetramer (Schell 1993).

3. Location and size of binding site of MJ-LysR protein.

The location and size of the MJ LysR binding site are consistent with those of other members in the LysR family, which are usually located close to the 5'-end of the gene encoding for the LysR protein with size varying from 26 bp to 54 bp (Schell 1993). The weak hypersensitive bands on both strands perhaps indicate conformational changes upon the binding of the protein. The conformational changes indicated by the weak bands may imply the requirement of an inducer in order to bring the more conformational changes that is needed to fully transcribe the adjacent divergent gene, but the identity and nature of inducer remain unclear to date.

To identify a possible consensus recognition motif for the binding of MJ-LysR protein, the protected sequences of Dnase I footprinting experiments were examined closely. The typical recognition sequences for most of the LysR-type transcriptional regulators are characterized by a T-N_{11}-A motif that is usually part of a larger dyadic sequence (Schell 1993). Such a typical recognition sequence, TTAGTTATTCTAA, is present within the protected region of the MJ-LysR protein. Furthermore, a TAGG$^{G}_{A}$ATTA sequence was found to be present twice on coding strand. The entire protected region by the MJ-LysR protein showed symmetry, although it is not perfect, with the $TN_{11}A$ recognition motif located at the center and two repeating sequences distributed on the both sides of the $TN_{11}A$ motif (Figure 2-17). Each repeating sequences might represent the binding site for a dimer.

4. Thermostability of biological macromolecules

After pre-heating at 94°C for 30 min, the MJ-LysR protein still maintained at least 75% of its binding activity as shown in the gel shift assay (Figure 2-8). This result seems inconsistent with the result that the protein lost its binding activity when incubating with DNA at 94°C in Figure 2-7. In fact, the DNA used in the gel mobility shift assay is naked DNA. In archaea, DNA adopts a unique topology containing positive supercoils (Forterre et al. 1996). Although the role of DNA overlinking in hyperthermophiles is still not clear (Forterre et al. 1996), at least in this case, our results indicate that the DNA topological state does contribute to the thermostability of DNA. A major challenge for DNA molecules in hyperthermophilic environments is

depurination followed by breaking of the phosphodiester bonds (Forterre et al. 1996). In an *in vitro* gel mobility shift experiment, the entire DNA is in a relaxed state, so it is extremely fragile to high temperatures. Moreover, the A+T (adenine/thymine) content in *M. jannaschii* is relatively high, about 57.4%. The high percentage of A+T makes the DNA molecule even more sensitive to heat. It was concluded from these observations that it is the DNA denaturing that caused the failure of the binding of MJ-LysR protein at high temperatures, not the denaturing of the protein. The protein itself maintained its binding activity at 94°C, a result consistent with the growing temperature range of *M. jannaschii.*

One of the most interesting findings in the field of biochemistry in the past few decades is that some proteins function at temperatures above that of boiling water. The basis for the thermostability of these proteins is not clear although some clues are emerging. Increased numbers of H-bonding and ionic bridges, hydrophobic effects, and minimizing the ration of surface/volume have been suggested as the possible determinants for the thermostability of these proteins. Among all possible explanations for increased thermostability in nature, increased H-bonding and/or an ionic interaction network are considered as dominant factors in thermal stability (Vogt & Argos 1997). Coincident with this conclusion, a statistically increased numbers of ionic bridges or ionic bridge networks have been found in many thermophilic and hyperthermophilic proteins (Ladenstein & Antranikian 1998). The MJ-LysR protein also falls into this category. It contains a higher proportion of charged amino acids (Table 2-3) compared

with mesophilic bacteria. This increased charge-charge interactions might be an important factor contributing to the thermostability of this protein.

5. Prediction of promoter elements in *M. jannaschii* and transcription regulation of MJ-LysR in *M. jannaschii.*

A comparison of sequences upstream of archaeal genes has revealed that two conserved sequence elements are important for archaeal transcription *in vitro*. They are Box A (TATA box) and Box B. The Box A (TATA box) that is located approximately 30 bp upstream of the transcription start site corresponds to the TATA box in the eukaryotic RNA polymerase II promoter and has general consensus sequences: 5'$^{T}_{A}$ $^{T}_{A}$TATATA-3'. The Box B contains the transcription start site and has a weak consensus: 5'-$^{T}_{C}$ $^{G}_{A}$-3', with the transcription starting at the purine residue (Palmer & Daniels 1995). A third element, which is designated BRE (transcription factor TFB recognition element), was found immediately upstream of the Box A. This is a purine-rich element comprised at least two adenines. A piece of DNA, 5'-$^{T}_{A}$ $^{T}_{A}$TATATA-3' preceded by a run of purines was used as a probe to search for matches on the coding strand and noncoding strand of the intergenic control region between the MJ-LysR protein and its upstream gene. The potential promoter region for the MJ *lysR*-type gene was located at position –51 to –41. It is interesting that this region overlaps with the 5'-end of the MJ-LysR protein-binding site. This overlapping provides some hint as to the mechanism of transcription regulation in *M. jannaschii*. As

described above, the typical LysR-type transcriptional regulator positively regulates expression of its target gene or genes while negatively regulating expression of its own structural genes. The binding of the basal transcription apparatus, TBP and TFB to the Box A (TATA box) and its upstream BRE sequence respectively, helps to recruit the RNA polymerase to the promoter site. The presence of the MJ-LysR protein, most likely functioning as a transcription repressor for its own gene expression, will block the binding of TBP and TFB and lead to a failure of RNA polymerase recruitment. Since the intergenic DNA between the *lysR*-type gene and its upstream gene has a length of 361 bp, which is much longer than those in the other LysR family members, the promoter for the upstream gene may not overlap with that of the MJ *lysR*-type gene as the overlapping promoter structures shown by other *Lys*R-type genes. However, we showed the evidence of DNA bending induced by the binding of the MJ LysR protein to its binding site. This bending could bring the transcription regulator, MJ LysR protein bound far away from the promoter, which controls the expression of the gene upstream from the MJ *lysR*-type gene, closer to or even in contact with RNA polymerase and/or other general transcription factors bound at the promoter region, thus, facilitating the transcription of its adjacent upstream gene. Therefore, we think that the MJ-LysR protein on the other hand might function as an activator to stimulate its upstream gene expression. In this way, the eukaryotic-like transcriptional machinery (TBP, TFB and archaeal RNA polymerase) and the eukaryotic promoter structure coexist or cofunction with the bacterial transcription regulator (MJ-LysR protein) to stimulate the transcription of the *lysR*-type genes in this archaeon.

Therefore, based on the available evidence in this study and the knowledge of the LysR family of transcriptional regulators/archaeal transcription basal machinery, a chimaeric mechanism for transcription regulation in *M. jannaschii* was proposed. The bacterial-like transcriptional regulator, the MJ-LysR protein, working with the eukaryotic-like basal transcription apparatus, namely, eukaryotic-like promoter elements, general transcription factors and archaeal RNA polymerase, forms a DNA-protein complex (transcription initial complex) that is similar but much simpler than that formed in eukaryotes. The necessary conformational changes in DNA and/or DNA-protein complex upon binding of proteins would be brought in and start the transcription.

Summary

In this study, I described the subcloning/expression of the putative *lysR*-type gene from *M. jannaschii*, purification of the gene product, and analysis of its DNA binding function *in vitro*. The protein encoded by the *lysR*-type gene from *M. jannaschii*, termed MJ-LysR protein, is homologous to the bacterial transcriptional regulators in the LysR family. I showed that the purified MJ-LysR protein bound to the intergenic control region specifically and selectively. The binding site was located by Dnase I footprinting experiments. The size and location of the binding site are consistent with the other LysR-type transcriptional regulators. The potential promoter element of this MJ *lysR* gene was predicted by sequence comparisons. By analyzing all these data, a chimaeric mechanism of transcription regulation in this archaeon is proposed: that is, this bacterial-like transcriptional regulator combines with the eukaryotic-like basal transcription apparatus to regulate transcription in *Methanococcus jannaschii*.

Table 2-1. Plasmids and bacterial strains

Name	Characteristics	Source
JM 109	*recA*1, *endA*1, *gyrA*96, *thi*, *hsdR*17(r_k^-, m_K^+), *supE*44, Δ(*lac-proAB)*, *relA*1, [F', *traD*36, *proAB*$^+$, *lacI*^{q}ZΔM15], λ^-	
	Cloning host strain	Promega
BL21(DE3)pLysS	F^-, *ompT*, *hsdS*$_B$, ($r_B^-m_B^-$), *dcm*, *gal*, (DE3), pLysS, Cm^r	
	Expression host strain	Promega
*p*ET-5b	Bacterial expression vector, Amp^r, origin of *E. coli*, T_7 transcription control region.	Promega
*p*AMBIJ48	Cloning vector with the *lysR*-type gene from *M. jannaschii.* Amp^r.	ATCC*

*American Type Culture Collection.

Table 2-2. Primers used in this study

Primer	Sequences	Usage
JK87	GTC CAC AGA ACA TAT GGA TCC* AAA AAT AAG TTA	Upstream primer for *lysR* gene
JK88	CCA CAG AAT TC#T TAA ACC TTT GTA ACA AA	Downstream primer for *lysR* gene
JK89	TAA TAC GAC TCA CTA TAG GG	Sequencing primer
JK82	CCA CAG GAT CCT ACA TGC TGG AAA GAT TA	Sequencing primer
JK90	GCC CAC ATA AAC CTC CAA CAT CTG	Upstream primer
JK91	AGT ATT TCT CAA GTG CTG ATA TGT	Downstream primer
JK92	GTG TAA GTT TTG GAA AGC AC	Upstream primer
JK93	CTT GCA ACT ATA AAT GTT TG	Downstream primer
JK94	GTA GTT TCG GGA AAT TAC GAC	Upstream primer
JK95	CCC TAT TAG AAT AAC TAA TA	Downstream primer
JK97	ATT CCA TTG GTA CTT AAA ATT CA	Upstream primer

*: the *Bam*HI cut site.

#: the *Eco*RI cut site.

Table 2-3. Charge amino acid content of some thermophilic and mesophilic LysR-type proteins. * denotes thermophilic organism. (Keller, personal communication)

	Charged Amino Acids				
Organism	Asp	Glu	Lys	Arg	Sum
**A. aeolicus*	18	21	32	14	85
**M. jannaschii*	17	24	28	13	82
**A. fulgidus*	17	23	28	13	81
B. Subtilis	16	25	24	13	78
B. cepacia	21	12	5	26	63
E. coli	12	22	5	26	64

Table 2-4. PCR yields for each reaction component at each level tested

	Level		
	A	B	C
primer	0-3274-3418	1279-1221-0	3744-0-2206
template	0-1279-3744	3274-1221-0	3419-0-2206
$MgCl_2$	0-0-2206	3274-1279-2206	3419-1221-3744
dNTPs	0-1221-2206	3274-0-3744	3419-1279-0

Each number represents amplification yields for reactions containing the component at concentration levels A, B, or C. PCR product yields were calculated as relative peak area.
The concentration of each component at level A, B, and C is listed as below:

	A	B	C
each primer (pmole)	7.5	15	30
DNA template (ng)	5	15	30
$MgCl_2$ (mM)	0.5	1.0	5.0
dNTPs (mM)	0.05	0.2	0.4

Table 2-5. Signal-to-noise ratios (SNL) for reaction components calculated using PCR product yields

	level		
	A	B	C
Primer	4.77	4.77	4.77
Template	4.77	4.77	4.77
$MgCl_2$	1.76	65.1	65.5
dNTPs	4.77	4.77	4.77

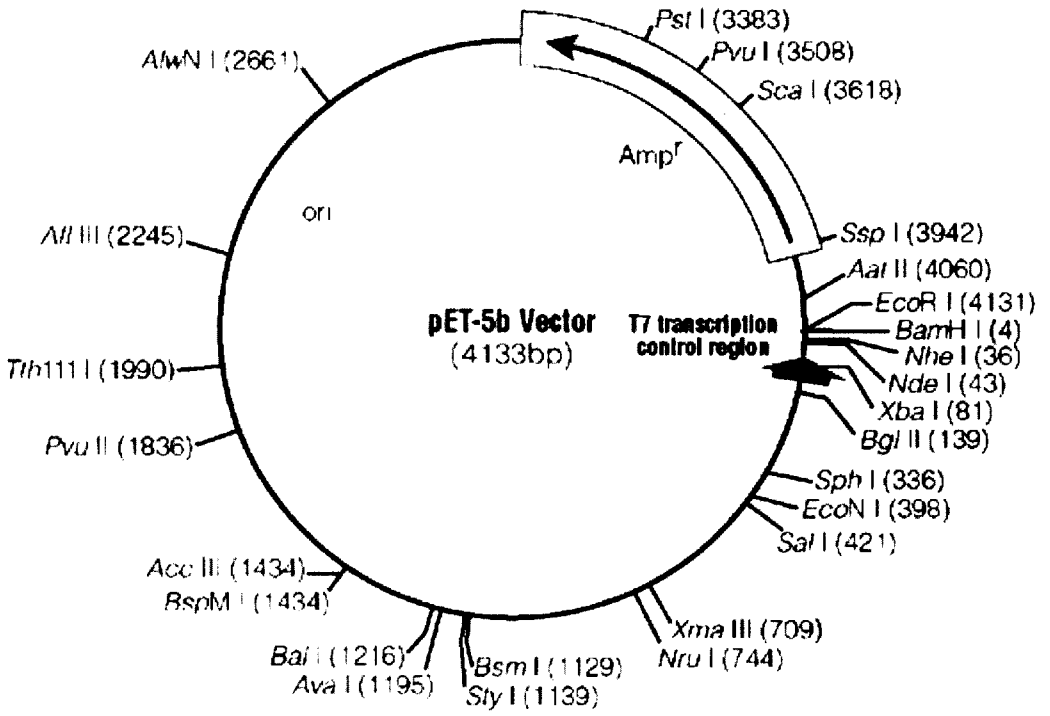


Figure 2-1. A map of pET-5b vector and sequence reference points. Locations of unique restriction enzyme cutting sites are indicated on the map.Amp^r, gene conferring ampicillin resistance in *E.coli*; ori, origin of replication in *E .coli.* The arrow within the Amp^r indicates the direction of transcription. The solid arrow, which encompasses the T7 promoter, shows the direction of transcription from the T7 promoter (pET-5b expression vectors technical manual).

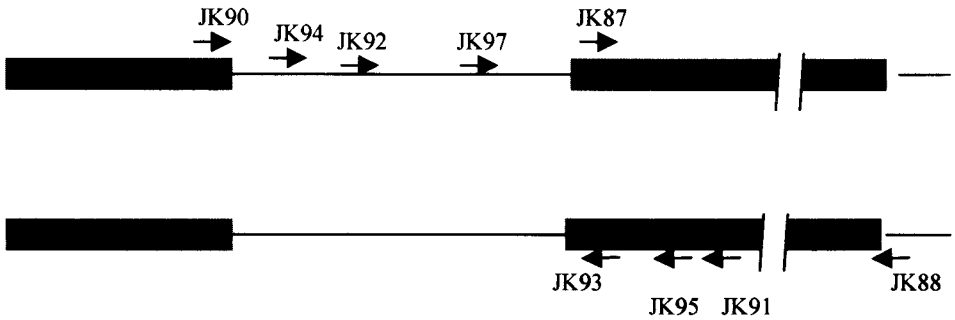


Figure 2-2. The relative positions of primers on double-stranded DNA

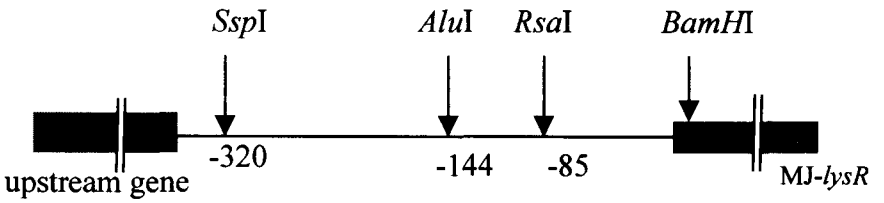


Figure 2-3. Positions of restriction enzyme cutting sites on the DNA fragment.

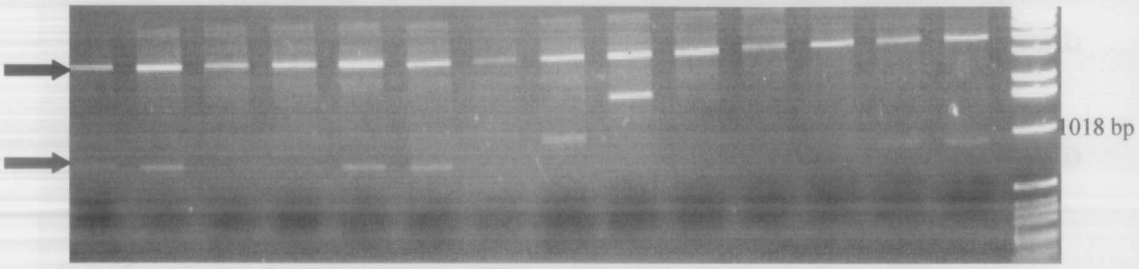


Figure 2-4. Recombinant plasmid *p*ET-5b/MJ*lysR* digested by *Bam*HI and *Eco*RI. The bottom arrow indicates the insert *lysR*-type gene from *M. jannaschii.* The top arrow indicates the *p*ET-5b vector.

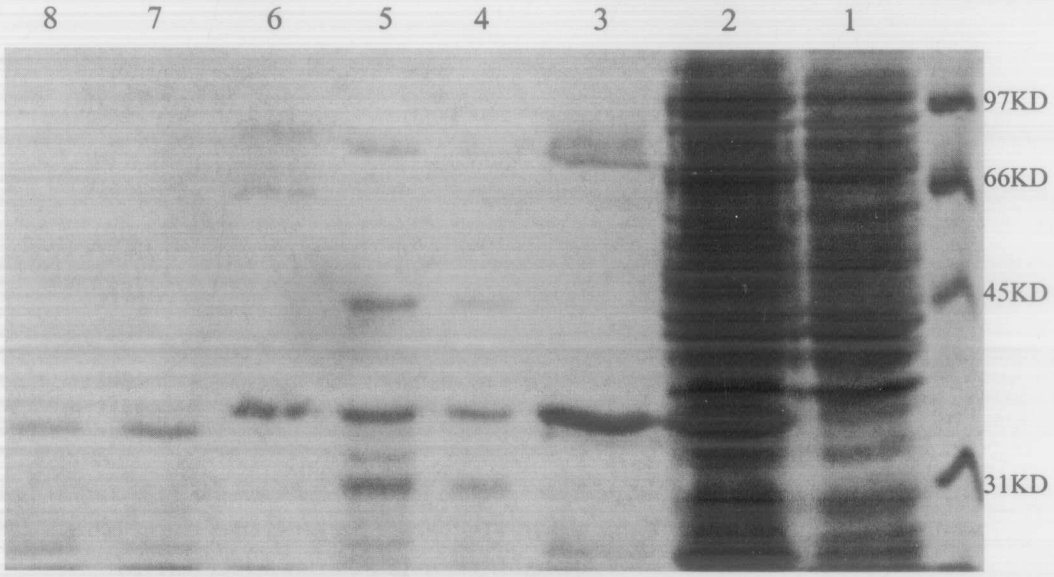


Figure 2-5. SDS-PAGE analysis of the MJ-LysR protein purification.
M: low range protein marker.
1: crude protein solution in uninduced *E. coli* cells.
2: crude protein solution in IPTG-induced *E. coli* cells.
3: crude protein solution after heating at 85˚C for 20 min.
4-8: fractions after size exclusion chromatography.

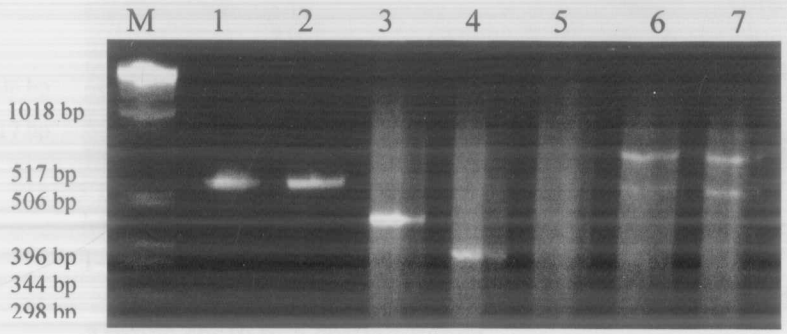


Figure 2-6. Binding specificity of the purified Mj-LysR protein
M:1kb DNA molecular weight ladder,
1: the intergenic DNA fragment between MJ lysR gene and its upstream gene amplified from pAMBIJ46 using primer JK90 and JK91.
2: the intergenic DNA fragment plus 30 ng BSA,
3, 4: the intergenic DNA fragment between MJ *lysR* gene and its upstream gene from *Burkholderia* cepacia plus 30 ng LysR protein from *M. jannaschii.*
5: LysR protein from *M. jannaschii.*
6, 7: the intergenic DNA fragment between lysR gene and its upstream gene from *M. jannaschii* plus 30 ng LysR protein from *M. jannaschii.*

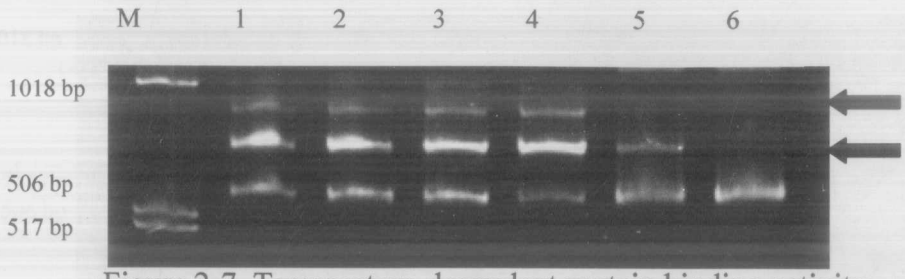


Figure 2-7. Temperature-dependent protein binding activity
M: 1kb DNA molecular weight ladder.
1-6:LysR protein from *M. jannaschii* interacts with the intergenic DNA fragment at 0°C, 25°C, 37°C, 64°C, 85°C and 94°C.
Arrows show two shifted bands

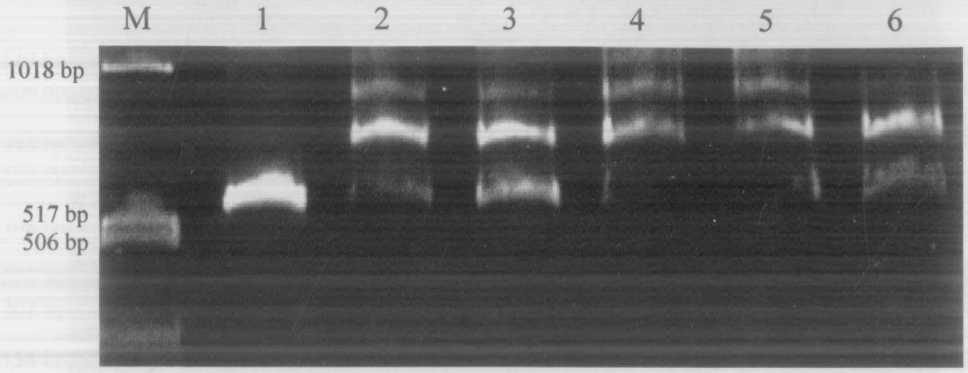


Figure 2-8. Thermal stability of MJ-LysR protein.
M: 1kb DNA molecular weight ladder.
1: 555 bp DNA intergenic fragment only.
2: 70 ng MJ-LysR protein preincubated on ice incubated with 555 bp DNA intergenic fragment at room temperature.
3: 35 ng MJ-LysR protein preincubated at 25°C incubated with 555 bp DNA intergenic fragment at room temperature.
4: 70 ng MJ-LysR protein preincubated at 64°C incubated with 555 bp DNA intergenic fragment at room temperature.
5: 70 ng MJ-LysR protein preincubated at 79°C incubated with 555 bp DNA intergenic fragment at room temperature.
6: 70 ng MJ-LysR protein preincubated at 94°C incubated with 555 bp DNA intergenic fragment at room temperature.

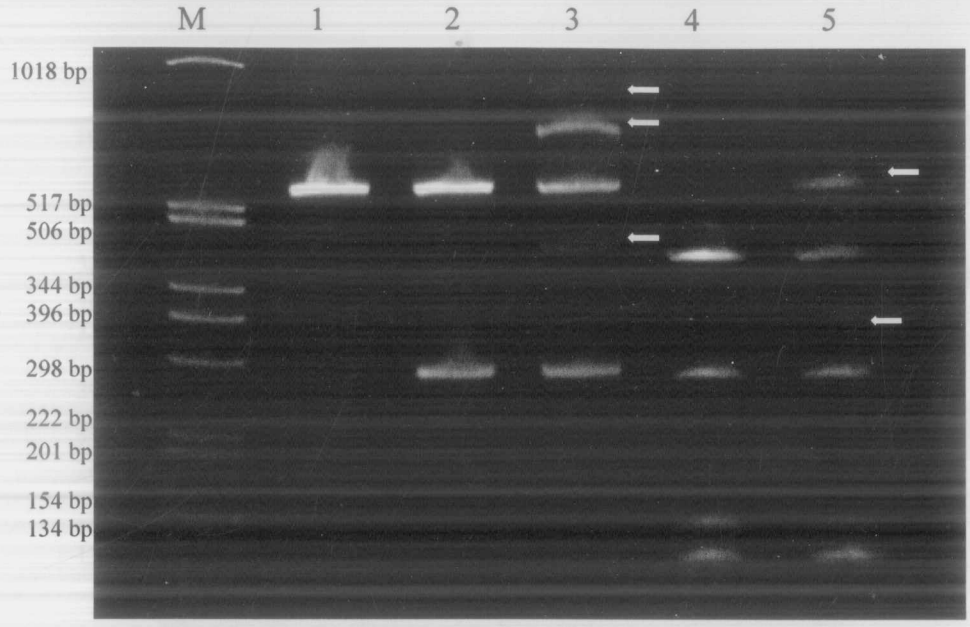


Figure 2-9. *Alu*I and *Alu*I+*Bam*HI digestions of the 555 bp intergenic DNA fragment.
M: 1kb DNA molecular ladder.
1: 555 bp intergenic DNA fragment.
2: 555 bp intergenic DNA fragment cut by *Alu*I (incomplete digestion). fast moving band contains two fragments: 278 bp and 277 bp.
3: 555 bp intergenic DNA fragment cut by *Alu*I incubated with purified MJ-LysR protein. The bottom arrow indicates shifted bands from one of fast moving DNA fragment. The top two arrows indicate shifted bands from uncut 555 bp DNA fragment.
4: 555 bp intergenic DNA fragment cut by *Alu*I and *Bam*HI: 130 bp, 147 bp, 278 bp DNA fragments and one incompletely cut band of size 425 bp.
5: 555 bp intergenic DNA fragment cut by *Alu*I and *Bam*HI incubated with purified MJ-LysR protein. The bottom arrow shows shifted band from 147 bp fragment. The top white arrow indicates shifted band from incompletely digested band.

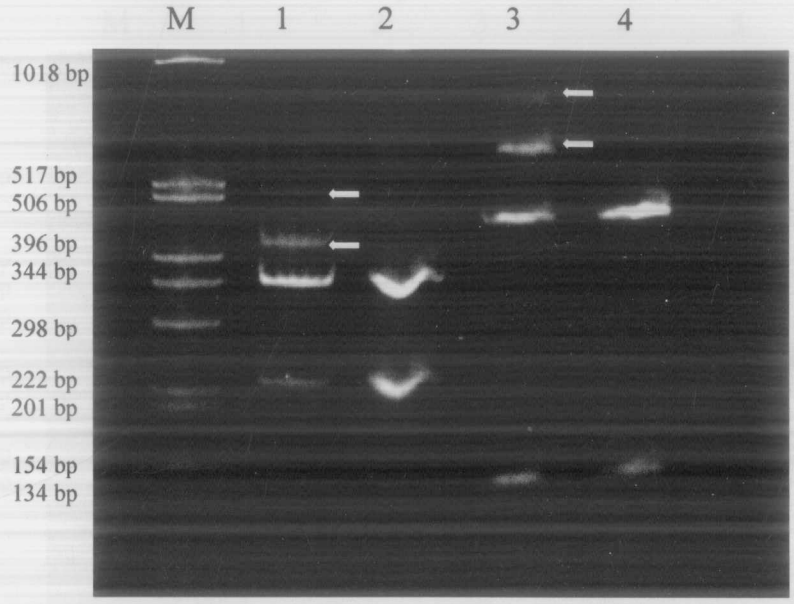


Figure 2-10. *Rsa*I and *Bam*HI digestions of the 555 bp intergenic DNA fragment
M: 1kb DNA molecular weight ladder.
1: 555 bp intergenic DNA fragment cut by *Rsa*I incubated with MJ-LysR protein. The arrows indicate shifted bands from 217 bp fragment.
2: 555 bp intergenic DNA fragment cut by *Rsa*I:338 bp and 217 bp fragments.
3: 555 bp intergenic DNA fragment cut by *Bam*HI incubated with MJ-LysR protein. The arrows indicate shifted bands from 425 bp fragment.
4: 555 bp intergenic DNA fragment cut by *BamH*I: 130 bp and 425 bp fragments

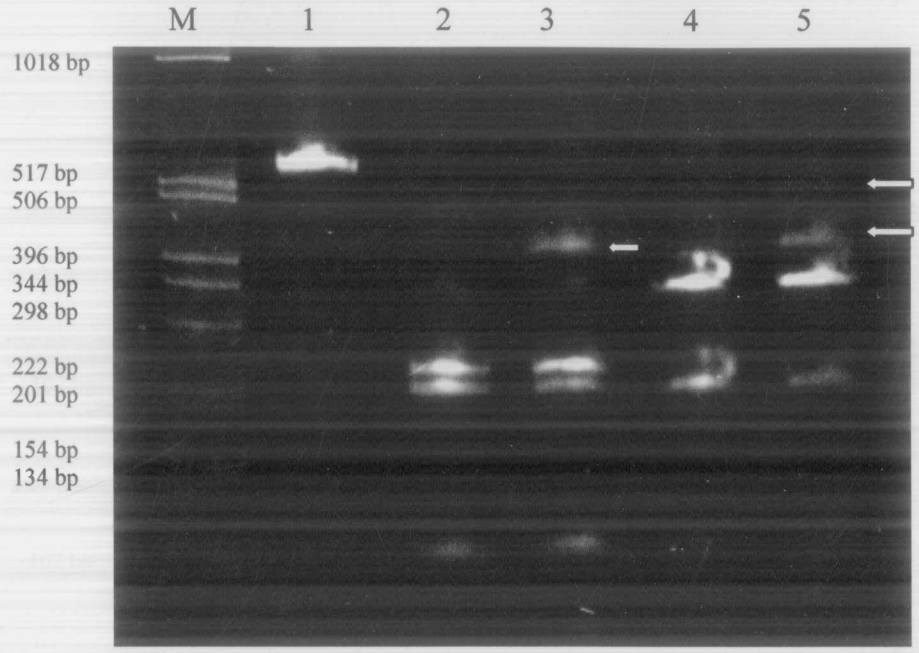


Figure 2-11. *Rsa*I and *Ssp*I digestions of the 555 bp intergenic DNA fragment.
M: 1kb DNA molecular weight ladder.
1: 555 bp intergenic DNA fragment only.
2: 555 bp intergenic DNA fragment cut by *Rsa*I and *Ssp*I: 102 bp, 217 bp and 236 bp fragments
3: 555 bp intergenic DNA fragment cut by *Rsa*I and *Ssp*I incubated with MJ-LysR protein. Arrow indicates shifted bands from 217 bp fragment
4: 555 bp intergenic DNA fragment cut by *Rsa*I only: 338 bp and 217 bp DNA fragments.
5: 555 bp intergenic DNA fragment cut by *Rsa*I incubated with MJ-LysR protein. Arrows indicate shifted bands from 217 bp fragment.

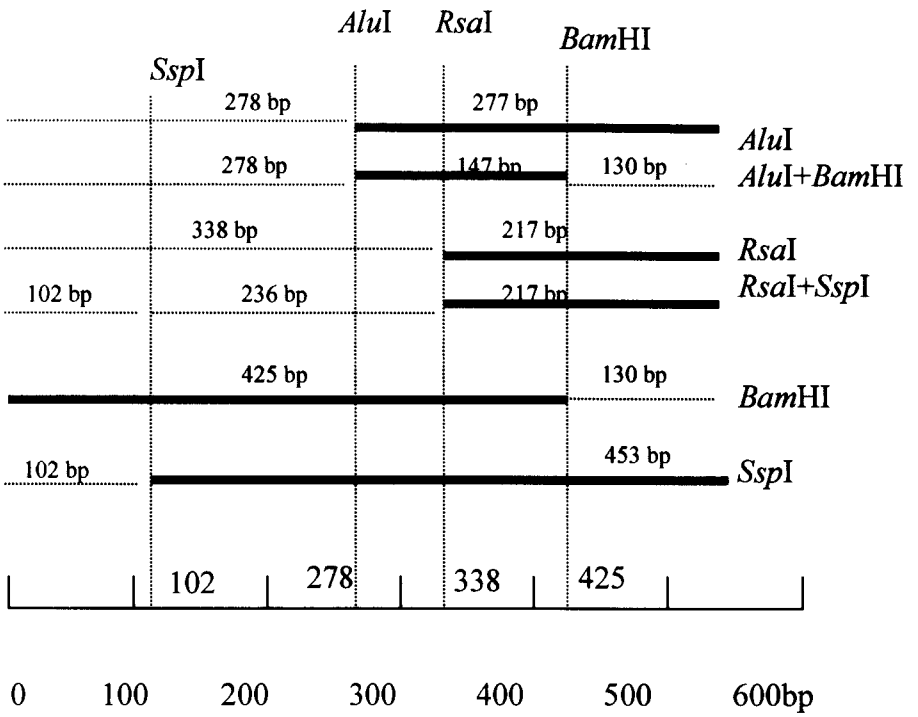


Figure 2-12. Gel mobility shift analyses of the 555 bp intergenic DNA fragment. Solid line represents the DNA fragment containing a MJ LysR binding site. Dotted line represents the DNA fragments not containing a MJ LysR binding site.

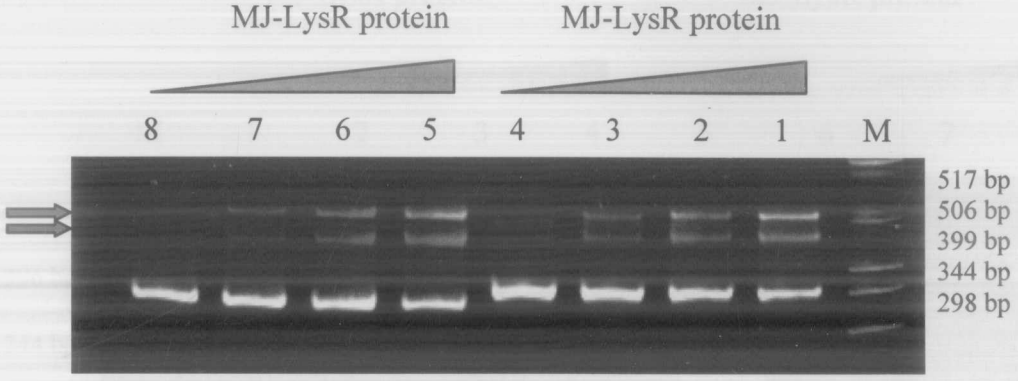


Figure 2-13. Estimation of the binding site of theMJ-LysR protein (I).
M: 1kb DNA molecular weight ladder.
8: 307 bp intergenicDNA fragment amplified by PCR using primer JK92/JK91.
4: 320 bp intergenic DNA fragment amplified by PCR using primer JK94/JK93.
7-5: 307 bp DNA fragment incubated with 10ng, 20ng, 40ng purified MJ-LysR protein.
3-1: 320 bp DNA fragment incubated with 10ng, 20ng, 40ng purified MJ-LysR protein.
Arrows indicate shifted bands.

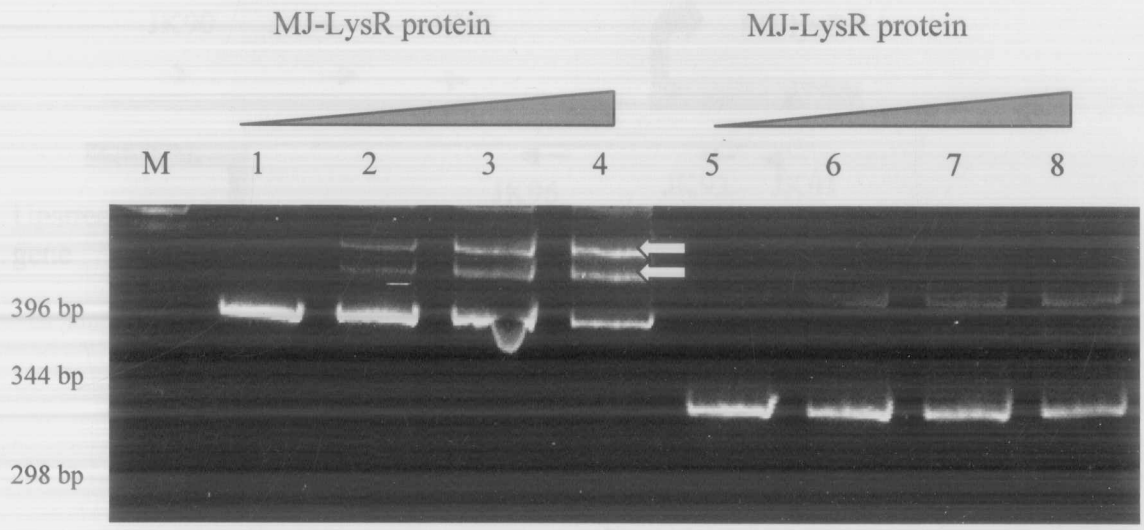


Figure 2-14. Estimation of the binding region of theMJ-LysR protein (II).
M: 1kb DNA molecular weight ladder.
1: 406 bp intergenic DNA fragment amplified by PCR using primer JK90/93.
5: 340 bp intergenic DNA fragment amplified by PCR using primer JK90/96.
2-4: 406 bp DNA fragment incubated with 10 ng, 20 ng, 40 ng MJ-LysR protein.
6-8: 340 bp DNA fragment incubated with 10 ng, 20 ng, 40 ng MJ-LysR protein.
Arrows indicate shifted bands.

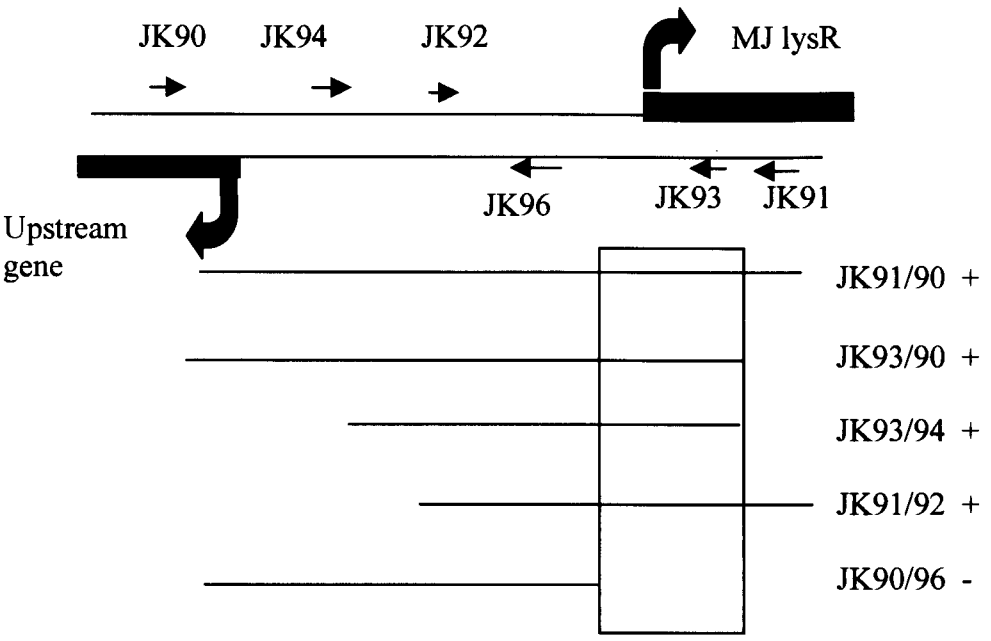


Figure 2-15. Summary of estimation of the MJ-LysR binding region.
+: the fragment containing a MJ LysR binding site; -: the fragment not containing a MJ LysR binding site.
The boxed region indicates a MJ LysR binding site.

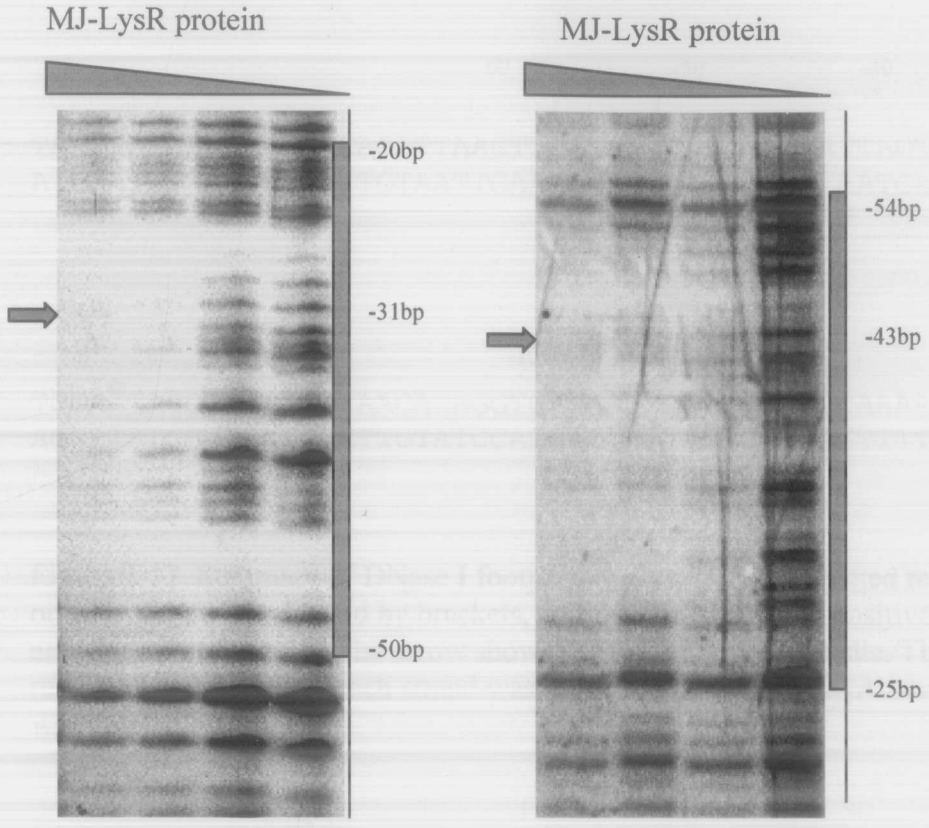


Figure 2-16. DNase I footprinting analyses of MJ-LysR protein binding to the control region of its regulated gene on the coding strand (A) and on the noncoding strand (B).
Arrows indicate hypersensitive bands on both strands

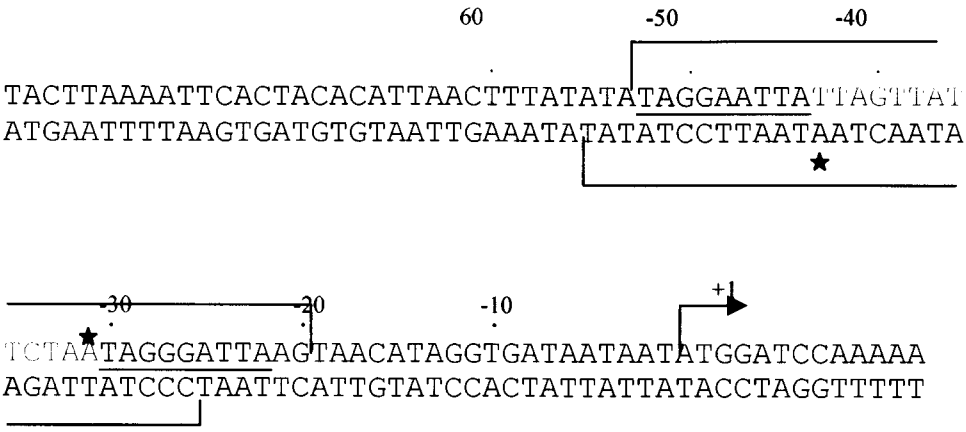


Figure 2-17. Summary of DNase I footprint analyses. The protected region on each strand is indicated by brackets, and the Dnase I hypersensitive sites are shown by asterisks. The arrow shows the translational start site. The repeating sequences on each strand was underlined and the $TN_{11}A$ motif was labeled by red letters.

```
GATTGTTACT TTGTTGTTTT CCATtctgct caccctgttt tttgaatttt gttaaaatct 60

                                  -295        -285
ttaatattta tatttagttg ggataattag tatataaatt taactttaag tagtttcggg 120
                                   ATATATTTaa a

aaattacgac tatttaaaac acaaaggaga tataaaaatc taatcctgta actgttgtca 180

tctacaacta atttaaatat atatgctttg agtgtaagtt ttggaaagca ctataaaaaa 240

gctattttct ctcaataatt tattatagat tttttatatt tatacttact attccattgg 300

                                     -51         -41
tacttaaaat tcactacaca ttaactttat atataggaat tattagttat tctaataggg 360
                                              [redacted]

attaagtaac ataggtgata ataatATGGA TCCAAAAATA AGTTATTTTC AAACATTTAT 420
[redacted]

AGTTGCAAGT AAAACAAAAA GTTTTTCTAA GGCAGCAAAA AGATTGGGAA TTACTCAAGG 480
```

Figure 2-18. Search for promoter elements on the coding and noncoding strand of the intergenic DNA region. The potential TATA box is underlined and the purine-rich BRE site is labeled by red letters on coding strand. The only possible promoter element forMJ *lysR*-type gene is located at –51 bp to –41 bp relative to the MJ LysR translation start site. This is the only one preceded by a run of purines. The MJ LysR protein binding site is shown by a black bar. On the complementary noncoding strand, the only potential promoter for the gene upstream from the MJ *lysR* gene is located at -285 bp to –295 bp relative to the MJ LysR translation start site.

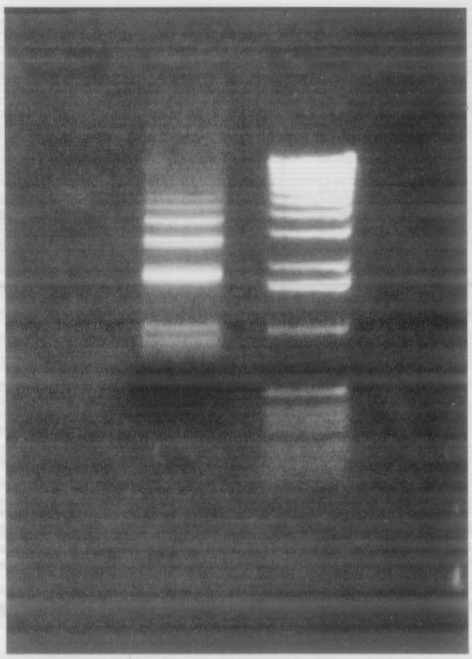


Figure 2-19. Ligation test on avalaibility of the sticky ends generated by *Bam*HI and *Eco*RI.

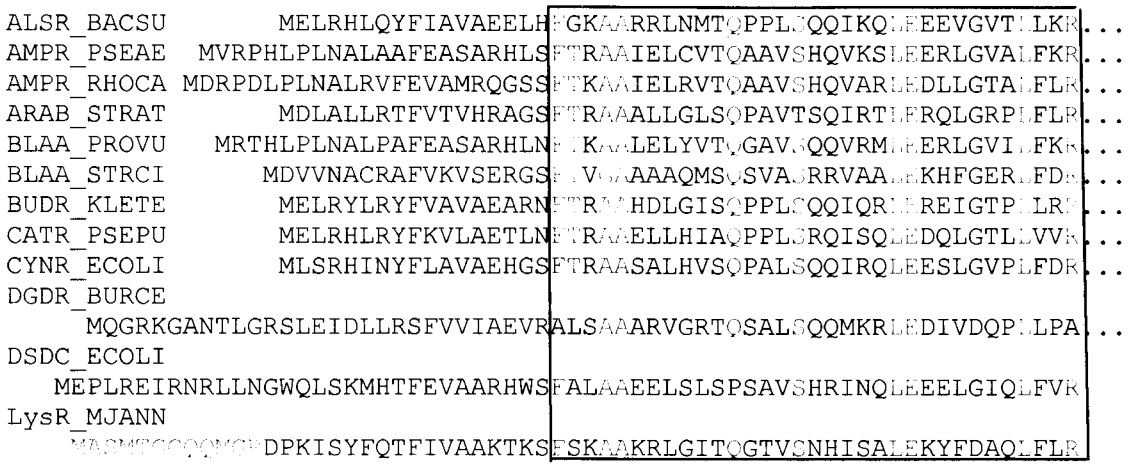


Figure 2-20. The N-terminal amino acids alignment of several LysR-type transcriptional regulators. DNA binding motif helix-turn-helix is boxed. The highly conserved amino acids are labeled red. The extra amino acids at N-terminal of the fusion MJ-LysR protein are labeled green.

Appendix I. An example of quantitative analysis of DNA molecule

After the gel was stained in a staining solution and rinsed with the deionize water for several time, it was analysis with AlphaImage 3.2 System. The bands on the gel were scanned; the area occupied by the band was reported. For example, the gel in Figure 2-9 was scanned, the areas occupied by the two bands in lane 2 are 2167 and 895 from top to bottom. The total area occupied is: 2167+895=3062. The total DNA in the reaction is 30 ng. So, the DNA present in the fast-moving band (the bottom band) is calculated as following: x is the amount of DNA present in the bottom band.

2165 : 895 == (30-x) : x

x = 8.9 ng

In lane 3, the area occupied by the very bottom band decreased from 895 to 706. The total area occupied by the total DNA is the sum of the area occupied by each band. In this case, it is

224+992+52+1108+193+706=3223

if let x be the amount of DNA present in the very bottom band, then x can be solved by the following equation:

706 : 3223 = x: 30

x= 6.6 ng.

In other word, 6.6 ng of DNA is present in the very bottom band.

Appendix II Row data of DNA quatitative analysis

All the area readings occupied by each band were scanned by AlphaImage 3.2 software.

The raw area data for Figure 2-9:

Lane 1	Lane 2	Lane 3	Lane 4	Lane 5
3087	2167	224		175
		992	1252	794
		1108		177
		193	583	557
	895	706	360	181
			598	564

The raw area reading for Figure 2-10:

Lane 1	Lane 2	Lane 3	Lane 4
181		353	
585		932	
1152	1986	1183	2032
350	589	570	679

The raw area reading for Figure 2-11

Lane 1	Lane 2	Lane 3	Lane 4	Lane 5
3123		141		172
		853		694
	167	161	1587	1775
	2376	2154	832	594
	1540	719		
	400	383		

References:

Aububel, F.M., Brent, R., Kinston, R.E., Moore, D.D., Smith, J.A. and Struhl, K. 1995. Current Protocols in Molecular biology (volume I). *John Wiley & Sons Inc.*

Bell, S.D., Cairns, S.S., Robson., R.L., and Jackson, S.P. 1999. Transcriptional regulation of an archaeal operon in vivo and vitro. *Mol. Cell* 4: 971-982.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*72: 248-254

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, K.D., et al. 1996. Complete genome sequence of the Methanogenic Archaeon, *Methanococcus jannaschii. Science* 273: 1058-1073.

Byerly, K.A., Urbanowski, M.L., and Stautfer, G.V. 1991. The MetR binding site in the *Salmonella typhimurium metH* gene: DNA sequence constraints on activation. *J. Bacteriol.* 173(11): 3547-3553.

Chung, C.T., Niemela, S.L., and Miller, R.H. 1989. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA*. 86: 2172-2175.

Cobb, B.D. and Clarkson, J.M. 1994. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. *Nucleic Acids Res.* 22(18): 3801-3805.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J. and Swanson, R.V. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392: 353-358.

Forterre, P., Bergerat, A., and Lopez-Garcia, P. 1996. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* 18: 237-248.

Gelfand, M.S., Koonin, E.V., and Mironov, A.A. 2000. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res.* 28(3): 695-705.

Goethals, K., Montagu, M.V., and Holsters, M. 1992. Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORF 571 reveal a common structure in promoters regulated by LysR-type proteins. *Prod. Natl. Acad. Sci. USA.* 89(5): 1646-1650.

Jones, W. 1983. *Arch.Microbiol.* 136: 256.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, K.D., Richardson, D.L.,Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Sill, S., Kirkness, E.F., Dougherty, B.A., Mckenney, K., Adams, M.D., Loftus, B., Venter, J.C., et al. 1997. The Complete Genome Sequence of the Hyperthermophilic, Sulphate Reducing Archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364-370.

Napoli, A., Van der Oost, J., Sensen, C.W., Charlebois, R.L., Rossi, M., and Ciaramella, M. 1999. An Lrk-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. *J. Bacteriol.* 181(5): 1474-1480.

Palmer, J.R. and Daniels, C.J. 1995. In vivo definition of an archaeal promoter. *J. Bacteriol.* 177(7): 1844-1849.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning. 2nd edition. *Cold Spring Harbor Laboratory Press*

Schell, M.A. and Poser, E.F. 1989. Demonstration, characterization, and mutational analysis of NahR protein binding to *nah* and *sal* promoter. *J. Bacteriol.* 171(2): 837-846.

Schell, M.A. 1993. Molecular biology of the LysR family of trnascriptional regulators. *Annu. Rev. Microbiol.* 47: 597-626.

Soppa, J. 1999. Transcription initiation in Archaea: facts, factors and future aspects. *Mol. Microbiol.* 31(5): 1295-1305.

Tang, X., Nakata, Y., Li, H, Zhang, M., Gao, H., Fujita, A., Sakatsume, O., Ohta, T., and Yokoyama, K. 1994. The optimization of prepations of competent cells for transformation of *E. coli. Nucleic Acids Res.* 22(14): 2857-2858.

Thomm, M. 1996. Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol. Rev.* 18: 159-171.

Tumbula, D. and Vhitman, W.B. 1999. Genetics of *Methanococcus*: possibilities for functional genomics in Archaea. *Mol. Microbiol.* 33(1): 1-7.

Tyrrell, R., Verschueren, K.HG., Dodson, E.J., Murshudov, G.N., Addy, C., and Wilkinson, A. 1997. The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Struc.* 5(8): 1017-1032.

Vogt, G. and Argos, P. 1997. Protein thermal stability: hydrogen bonds or internal packing? *Fold. & Des.* 2: S40-S60.

Chapter 3

Effects of Different Amino Acids on *in Vitro* DNA Binding Activity of DgdR Protein, a LysR-type Protein from *Burkholderia cepacia*

Abstract

The DgdR protein is a LysR-type transcriptional regulator from *Burkholderia cepacia*. It negatively regulates the expression of its divergent gene, *dgdA*. The *dgdA* gene encodes dialkylglycine decarboxylase that is responsible for decarboxylation and transamination in dialkylglycine metabolism. The amino acid, 2-methylalanine, is both the inducer for the DgdR regulated *dgdA* gene expression and substrate for decarboxylation. Eleven amino acids that have structures similar to 2-methylalanine were chosen to test their effects on the binding of the DgdR protein to its operator site. Among those amino acids tested, only 2-methylalanine (2MA), 1-aminocyclopentane-1-carboxylic acid (1A1CA), S-2-aminobutanoic acid (S2AB), RS-isovaline, and 2-trifluoromethyl-2-aminobutanoic acid generated the measurable band shifting in a gel mobility shift assay. D- or L-alanine, D- or L-norvaline, 2,2-diethyl glycine, and 2-trifluoromethylalanine did not cause any measurable band shifting. By closely examining the structures of these molecules, it was concluded that both alkyl side chain size and hydrophobicity are important factors for the inducer recognition and binding.

Introduction

The *dgdR-dgdA* gene from *Burkholderia cepacia* was first isolated and cloned in Keller's laboratory (Keller et al. 1990, 1994). The *dgdR* gene encodes a transcription repressor, designated DgdR, which belongs to the LysR-type transcriptional regulator family. The DgdR protein negatively regulates the expression of its divergent gene, *dgdA*. The *dgdA* gene encodes a vitamin B-6-dependent 2,2-dialkylglycine decarboxylase (EC4.1.1.64). These two genes, *dgdR* and *dgdA*, control the breakdown of 2-methylalanine in *B. cepacia.* The genes are separated divergently by a 78-base pair region, which contains sequences likely to comprise the promoters for both genes. The binding site for the DgdR protein is located from position -8 to -70 relative to the *dgdA* transcription start point (Allen-Daley et al. in preparation). The protected region of the DgdR covered the entire *dgdA* promoter and the adjacent 5'-terminus of the *dgdR* gene (Allen-Daley et al. in preparation). The original results (Keller et al. 1990) showed that D or L-alanine did not induce the conformational changes on the DNA-protein complex, but 2-methylalanine did. Therefore, we wished to further confirm this result by an *in vitro* binding assay.

In the studies presented here, eleven amino acids that have structures similar to that of 2-methylalanine were chosen to test how these chemicals affect the binding of the DgdR protein to the DNA fragment containing the promoter region of the *dgdA* gene. Our results suggest that the size and hydrophobicity of the side chains of these molecules strongly affect formation of the DNA–protein complex that is the active form for transcription.

Materials and Methods

1. Plasmids, bacterial strains, and growth conditions

All plasmids and bacterial strains used in this study are described in Table 3-1. All bacterial strains were grown in Luria-Bertani (LB) medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, H_2O, 1 L) at 37°C with gentle shaking or on LB plates (10 g tryptone, 5 g yeast extract, 10 g of NaCl, 15 g of agar, H_2O, 1 L). When necessary, ampicillin was added to a final concentration of 80 μg/ml. All water used was purified to >10 MΩ cm using a Milli Q water purification system (Millipore).

2. Primers and structures of the amino acids used in this study

The primers JK5-20 (5'-AGCTCTCCGGATCCAAGCTT-3') and JK55 (5'-CTCATCTCCCCCGAGGTGAA-3') were used to amplify a 594-bp DNA segment containing the 5'-end of the *dgdA* gene, the 5'-end of the *dgdR* gene, and the entire intergenic gene region (Figure 3-1). The structures of the amino acids used are compiled in Table 3-2.

3. Plasmid isolation and purification

The procedures for plasmid isolation and purification are the same as described in section 2 of chapter 2 using Wizard™ Mini DNA purification kit.

4. DgdR protein purification

The DgdR protein purification followed the purification steps developed in Keller's laboratory (Allen-Daley et al. in preparation). A 25 ml portion of overnight culture of the JM109 *E. coli* cells carrying pSB46 was used to inoculate 1 L of LB medium containing 80 μg/ml ampicillin and 8 μM IPTG (Isopropyl-β-D-thiolgalactoside) to an initial $A_{600} = 0.25$. After growing at 37°C with gentle shaking for 6 h, the culture was collected by centrifugation. The pellets were resuspended in 35 ml of cold extraction buffer (30 mM Tris, 20 mM KCl, 1mM EDTA, pH 8.0) and sonicated. After centrifugation (Sorvall RC-5B, GSA head, 20 min, 4°C, 5000 rpm), the supernatant was collected and diluted to 10 mg protein/ml using the same extraction buffer. Polyethyleneimine and NaCl were added to a final concentration of 37 μl/ml and 0.20 M, respectively. The resulting precipitate was collected by centrifugation (Sorvall RC-5B, SS34 head, 30 min, 4°C, 6000 rpm). The pellets were dissolved in 30 ml of buffer (1 M NaCl, 30 mM tris, 20 mM KCl, 1 mM EDTA, 5% glycerol), and then centrifuged again as above. The supernatant was brought to 1.25 M ammonium sulfate with a cold 3.8 M stock solution. The precipitate was collected by centrifugation and resuspended in buffer (30 mM tris, 1 mM EDTA, 0.5M KCl, 5% glycerol, pH 8.0) to a final protein concentration of 1 mg/ml. One and half ml of this solution was loaded onto the top of a 1.5 x 42 cm BioGel A-1.5 m column equilibrated with the same buffer. After chromatography, fractions were assayed for DNA binding activity using a gel mobility shift assay. The fractions with the highest specific activity were combined and stored at 4°C.

5. Amplification of the 594 bp DNA fragment by PCR

The polymerase chain reaction (PCR) was used to amplify the 594 bp DNA fragment from the plasmid pGEM7Z14/6a. The reaction was carried out in a 50 µl volume containing 1x *Taq* polymerase buffer, 5 U of *Taq* DNA polymerase, 200 µM 4 dNTPs, 1.5 mM of $MgCl_2$, 1 µM of each primer, 30 ng of pGEM7Z14/6a as a template, and water. The reaction conditions were 30 cycles of 97°C/45 s, 55°C/45 s, and 72°C/45 s on Coy TempCycler (Model 50/60, Coy Laboratory Products Inc). The PCR products were purified with StrataClean™ resin and Wizard™ PCR Preps DNA purification kit (Promega) as described in chapter 2 section 5.

6. Gel mobility shift assays

The binding activity of the DgdR protein to the intergenic DNA fragment was studied by gel mobility shift assays. Each reaction contained 30 ng of the 594 bp DNA, 2 µl 5x assay buffer (100 mM Tris at pH 7.5, 25 mM NaOAc, 25% (v/v) glycerol, and 0.005% BSA), 70 ng of the purified DgdR protein, and water to a final volume of 10 µl. After incubating at room temperature for 5 min, 1 µl of dye solution (40% sucrose and 0.1% bromophenol blue) was added to the each reaction. The mixtures were loaded onto a 4.5% polyacrylamide non-denaturing gel (29.2:0.8 acrylamide to bisacrylamide in 50 mM tris, 380 mM glycine, 2 mM EDTA, and 2.5% glycerol, pH8.0) pre-electrophorized at 97 V for 30 min. The electrophoresis was carried out at 97 V for 1 h. Following the electrophoresis, the gel was stained in 100 ml of a staining solution (10 µl of concentrated SYBR green in 100 ml of 1xTAE buffer or 40 µl of 0.55mg/ml

ethidium bromide in 100 ml of 1xTAE buffer) for 20 min. The stained gel was rinsed several times with deionized water and observed under UV-light with IS1000 Digital Image System (AlphaImager Version 3.2, Alpha Innotech Corporation).

To test the effects of the amino acids on the DNA binding ability of the DgdR protein, certain amounts of 2-methylalanine or other amino acid stock solutions were added to the reaction mixture to a final concentration of 0 mM, 0.1 mM, 0.2 mM, 0.5 mM, 1mM, 5 mM, 10 mM, or 17.5 mM. The mixtures were incubated and analyzed the same as described above.

To ensure the presence of 2-methylalanine during electrophoresis, a special comparison gel was cast with one half the gel (panel A) containing 10 mM 2-methylalanine incorporated in the gel buffer itself and the other half (panel B) without 2-methylalanine. The mixture for the binding reaction was prepared as described above except 2-methylalanine was not present in the reaction mixture and the final concentration of the DgdR protein in the reaction was 30 nM in lane 2 of panel A and panel B, 60 nM in lane 3 of panel A and lane B in Figure 3-5. The rest of the reaction conditions and electrophoresis conditions were the same as above.

Results

Among the 11 amino acids including 2-methylalanine, only 2-methylalanine, RS-isovaline, L-2-aminobutanoic acid, 1-aminocyclopentane carboxylic acid and 2-trifluoromethyl-2-aminobutanoic acid caused the measurable changes in the pattern of the bands in terms of mobility and proportion (Figure 3-2A and Figure 3-4) while the

others, D or L-alanine, D or L-norvaline, 2-trifluoromethylalanine, and 2,2-diethyl glycine, did not cause any measurable changes (Figure 3-3). The intensity of shifted bands in Figure 3-3 remained constant. The quantitative analyses of the effects of the amino acids were demonstrated in Figure 3-2B. With the increasing of amino acid concentration, the amount of free DNA (f) remained unchanged. However, more fast-moving DgdR-DNA complexes (d) were converted to the slow-moving DgdR-DNA complexes (t) and the mobility of the slow-moving band increased. The conversion indicates that the inducer molecules favor the formation of complexes of DNA and protein tetramer and in the meantime introduce some conformational changes in the complexes that lead to the change in band mobility. It is also interesting to note that unlike the other amino acids, 2-trifluoromethyl-2-amimobutanoic acid only induce the conversion at high amino acid concentration (> 10 mM).

When 2-methylalanine was incorporated in the gel instead of in the reaction mixture, the band shifting was even more obvious (Figure 3-5). On the left side of the gel in Figure 3-5 with the presence of 2-methylalanie, the mobility of the slow-moving band was significantly increased. With the increasing of DgdR protein, the proportion of the slow-moving band increased compared to what on the right side of the gel without the presence of 2-methylalanine.

Discussion

To test the effects of the amino acids with similar structure to 2-methylalanine on the DNA binding activity of DgdR protein, the inducer with different concentrations was

added to the reaction system. In the gel mobility shift assay without these inducer molecules as shown in the lanes with no added the amino acids tested in Figure 3-2, two bands with different mobility were generated. The fast-moving band was thought to be complex of DNA and DgdR dimer while the band with the slow migration was complex of DNA and DgdR tetramer (Schell 1993). With the presence of increasing amounts of these amino acids, the proportion of the fast-moving band decreased while the proportion and the mobility of the slow-moving band increased and the amount of the free DNA remained unchanged. It is obvious that upon binding of these amino acids, the more fast-moving complexes were converted to the slow-moving complexes. This indicates that the slow-moving complex (DNA-protein tetramer) might be the transcription-favored form. The changes in the mobility of the slow-moving band (Figure 3-2 and Figure 3-5) suggest that the inducer molecules might facilitate the conformational changes in the DNA-protein tetramer complex. These conformational changes are necessary for the DNA-protein complex to be maintained in the induced state or active form for transcription and important for turning on the *dgdA* gene expression. On the other hand, D- and L-alanine, D- and L-norvaline, 2-trifluoromethylalanine, and 2,2-diethyl glycine did not bring any changes in terms of the band position and intensity (Figure 3-4). All these results are consistent with in vivo *xylE* reporter gene expression experiments as shown in Table 3-3 (Allen-Daley et al. in preparation). In the *in vivo xylE* report gene expression experiments, the *xylE* reporter gene for catechol 2,3-dioxygenase was inserted downstream of the *dgdA* promoter, thus the *dgdA* gene expression could be monitored by the catechol dioxygenase assays. As

shown in Table 3-3, with the presence of 2-methylalanine, S-isovaline, 1-aminocyclopentane carboxylic acid, and S-2-aminobutanoic acid, the expression of the *xylE* gene is much higher than in the control. There is no significant difference between the control and the group with the presence of D-, or L-alanine, D- or L-valine, and D- or L-norvaline.

In addition, the results from this study demonstrated the potential of the *dgdR-dgdA* gene expression system as a gene switch system. The DgdR protein, a repressor of the *dgdA* gene expression, can serve as a switch in this system. The switch is at "off" position without the presence of the inducer molecules. These inducer molecules are the chemicals that can make band shifting in this study, such as 2-methylalanine shown in Figure 3-2. Upon the binding of the inducer molecule, the conformational changes and the interactions among DgdR protein, inducer, and the promoter region that are necessary for transcribing the gene were brought in. Therefore, the *dgdA* gene expression regulated by the DgdR protein could be turned on. The expression of the *dgdA* gene could be shut off at any time by withdrawing the chemicals as well.

By closely examining the structures of these 11 amino acids, it was concluded that both alkyl side chain size and hydrophobicity are important factors for inducer and substrate recognition and binding. Two-methylalanine, 1-aminocyclopentane-1-carboxylic acid, S-2-aminobutanoic acid, and RS-isovaline share similar structures and have similar molecular sizes while the other amino acids that did not introduce the changes in the fast-moving and slow-moving bands have either the larger or smaller *pro-S* or *pro-R* groups. Hydrophobicity is another factor. As

seen in Table 3-2, the amino acids having apolar groups being present at *pro-S* and *pro-R* positions are more likely to introduce the changes in intensity and mobility of the bands in a gel mobility shift assay, for example, 2-methylalanine, isovaline, 1-aminocyclopentane-1-carboxylic acid. The D- and L-alanine have a H^+ group either at *pro-R* or at *pro-S* position. This might decrease the hydrophobic interactions that are necessary for holding the inducer in the right position for the reaction. In the meantime, the smaller size of the H^+ group could be another reason contributing to the failure of the inducer binding as well. On the other hand, the amino acids having long *pro-R* or *pro-S* chains, such as D- and L-norvaline, might interfere with the binding of inducer due to the limiting space of the binding pocket, thus, led to the unchanged shifted band in Figure 3-3. These two amino acids failed to stimulate the gene expression in an *in vivo* gene expression as well (Table 3-3).

Although a H^+ group was present in the *pro-R* position of S-2-aminobutanoic acid, the $-CH_2CH_3$ group at the *pro-S* position might compensate for the loss of the hydrophobic interaction due to the presence of the H^+. This might the reason that S-2-aminobutanoic acid could introduce the changes in the band-shifting pattern (Figure 3-2A), the changes were not obvious as the changes introduced by 2-methylalanine and isovaline. The 2-trifluoromethylalanine has a very similar structure to 2-methylalanine except a $-CF_3$ group at the *pro-R* position instead of a $-CH_3$ group. The $-CF_3$ group is about 10% larger than $-CH_3$ group. This might be the reason that 2-trifluoromethylalanine did not bring any changes in the shifted bands. Like 2-trifluoromethylalanine, 2-trifluoromethyl-2-aminobutanoic acid has the $-CF_3$ group at

the *pro-S* position, but has $-CH_2CH_3$ group at the *pro-R* position. Why this amino acid affect band shifting behaviors while 2-trifluofomethylalanine does not cannot be explained here. Moreover, it remains unknown why 2-trifluoromethyl-2-aminobutanoic acid induced the band shifting only at high concentration. It might be related to the zwitterion concentration of 2-trifluoromethyl-2-aminobutanoic acid at physiological pH range.

Therefore, it was concluded from the above observations that the hydrophobic interactions at both *pro-R* and *pro-S* positions were necessary for holding the inducer in the appropriate position to facilitate the formation of DNA-protein complex and introduce the conformational changes in the complex. These conformational changes are necessary to stimulate gene expression. At the same time, the size of the inducer molecule is another factor that affects the binding of the inducer, thus affecting the induction of gene expression.

When 2-methylalanine was incorporated in the gel during electrophoresis, the band shifting was even more obvious as shown in Figure 3-5. On the left side of the gel with the presence of 2-methylalanine, the mobility of the slow-moving band was significantly increased. With the increasing of the DgdR protein, the proportion of the slow-moving band increased compared to the gel without 2-methylalanine. This experiment further indicated that 2-methylalanine did introduce some changes in the DNA-protein complex. These changes could come from the conformational changes due to the binding of the inducer. Comparing this result with the DNA footprint results,

it is obvious that the change is from shrinking of the binding site that frees the promter region of the *dgdA* gene.

In addition, from Figure 3-5, it is interesting to note the change of K_D (equilibrium dissociation constant) with the presence of the inducer molecule, 2-methylalanine. Normally, K_D is calculated as the total protein concentration that allows half-maximal DNA binding (Chugani et al. 1998). Thirty ng of DNA fragment labeled as a1 and b1 in Figure 3-5 was used in all binding reactions. When the DNA were incubated with 30 nM of DgdR protein, the free DNA in both panel A and B were 15 ng and 12 ng, labeled by a2 and b2, respectively (quantitative analyses were done using alphaImage 3.2 system). In other words, without the presence of 2-methylalanine in panel B of the gel, when the concentration of DgdR protein is 30 nM, 18 ng of total DNA was bound by the protein. Thus, K_D should be smaller than 30 nM. With the presence of 2-methylalanine, K_D is about 30 nM because this is the concentration of DgdR protein that made about half of the total DNA being bound by DgdR protein as shown in Table 3-4. It was concluded from here that K_D was increased due to the presence of the inducer, 2-methylalanine. The large value of K_D indicated the weaker-binding of the DgdR protein to its binding site with the presence of inducer, 2-methylalanine. This result is consistent with those from DNA footprinting assays that conformation of protein-DNA complex formed by DgdR protein and its binding site was altered upon addition of 2-methylalanine by freeing the –10 region. The increased value of K_D, on the other hand, provides an indirect evidence for DgdR protein functioning as a repressor. Without the absence of inducer molecule, DgdR binds to the *dgdA* promoter

region very tightly. The tight-binding of protein precludes the possibility of the recruitment of RNA polymerase. With the presence of the inducer molecule, the binding of DgdR protein becomes weak and leads to the conformational changes of protein-DNA complex that are necessary for inducing a gene expression.

Table 3-1. Bacterial strains and plasmids

Strain or plasmid	Genotype, phenotype, or properties	Source
E.coli JM 109	*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1λ⁻ Δ(lac-proAB)*	Promega
pGEM7Z14/6a	amp^r, pGEM7Z14 with 989-bp including 565-bp of 3' end of *dgdR* gene, deleted	this Lab
pSB46	amp^r, pBTac1 with *dgdR* gene inserted downstream from the *tac* promoter	this Lab

Table 3-2. Structures of amino acids used in this study

Names	α-substituents	
	pro-R	*pro-S*
2-methylalanine	CH_3-	-CH_3
S-2-aminobutanic acid	H-	-CH_2CH_3
1-aminocyclopentane-1-carboxylic acid	-CH_2CH_2-	-CH_2CH_2-
R-isovaline	CH_3CH_2-	CH_3-
S-isovaline	CH_3-	-CH_2CH_3
2-trifluoromethyl-2-aminobutanoic acid	CH_3CH_2-	-CF_3
L-alanine	H-	-CH_3
D-alanine	CH_3-	H-
L-norvaline	H-	$CH_3CH_2CH_2$-
D-norvaline	$CH_3CH_2CH_2$-	H-
2-trifluoromethylalanine	-CF_3	-CH_3
Diethyl glycine	CH_3CH_2-	CH_3CH_2-

Table 3-3. Catechol 2, 3-dioxygenase productions in *E. coli* JM109/pUCX11b treated with various amino acids (Allen-Daley et al. in preparation).

The upper group is inducers; the lower group non-inducers. Growth at 37°C in LB medium supplemented with 10 mM amino acid.

Amino acids	Specific activity (U/mg)*
2-methylalanine	361
S-isovaline	110
1-aminocyclopentane c.a.	88
S-2-aminobutanoic acid	71
D-norvaline	2
L-norvaline	5
D-alanine	6
L-alanine	18
D-valine	4
L-valine	7
none	2

* one unit = 1nmole α-hydroxymuconic ε-semialdehyde min^{-1} at pH 7.5,25°C.

c.a. = carboxylic acid.

Table 3-4. Change of the equilibrium dissociation constant (K_D) of DgdR protein upon binding of inducer, 2-methylalanine (2MA).

	free DNA without presence of 2MA (ng)	free DNA with presence of 2MA (ng)
0 nM DgdR	30 ng (a1 in Figure 3-5)	30 ng (b1)
30 nM DgdR	~13 ng (a2)	~15 ng (b2)
60 nM DgdR	~8.9 ng (a3)	~11.1 ng (b3)
K_D	<30 nM	~30 nM

K_D is defined as the total protein concentration that allows half-maximal DNA binding.

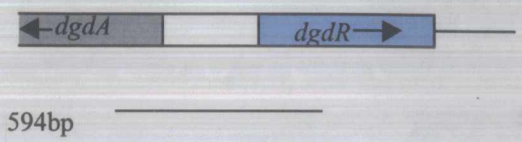


Figure 3-1. The diagram of the divergent *dgdR* and *dgdA gene* and the 594 bp DNA fragment used in gel mobility shift assays.

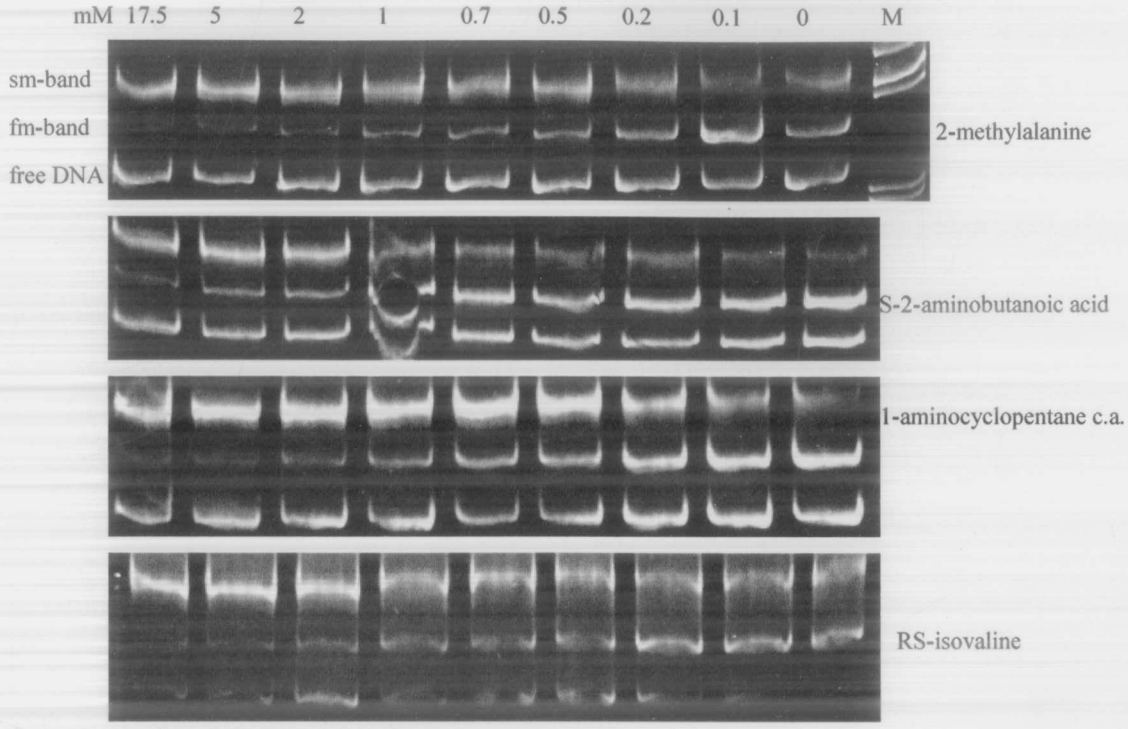
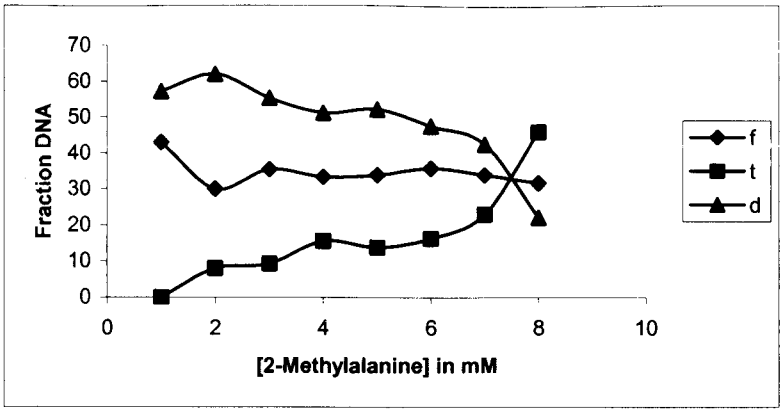
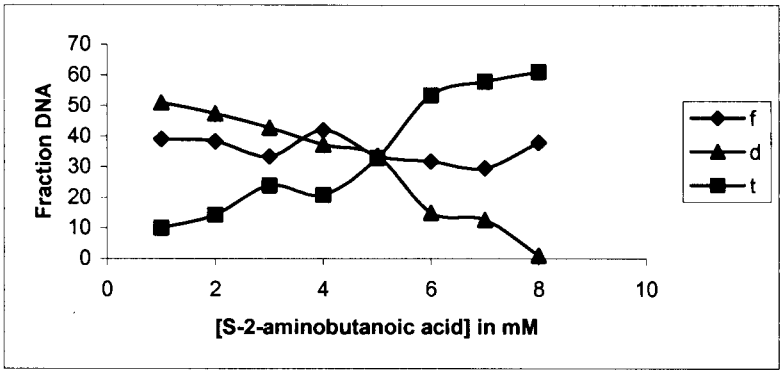
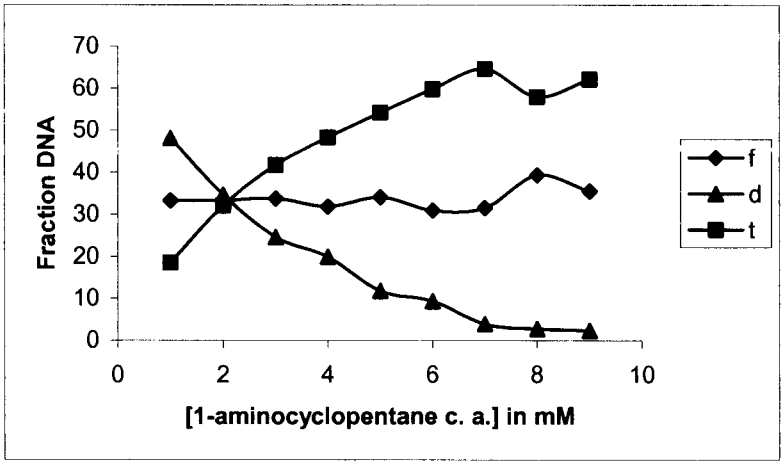


Figure 3-2A. Effects of different amino acids on DgdR-DNA complex formation (I): Amino acids have effects on DgdR-DNA complex. These chemicals increase the proportion of DNA in fm-bands and decrease the distance between fm-bands and sm-bands.


Fraction DNA
70
60
50
40
30
20
10
0
0
2
4
6
8
10
[2-Methylalanine] in mM
f
t
d


Fraction DNA
70
60
50
40
30
20
10
0
0
2
4
6
8
10
[S-2-aminobutanoic acid] in mM
f
d
t


Fraction DNA
70
60
50
40
30
20
10
0
0
2
4
6
8
10
[1-aminocyclopentane c. a.] in mM
f
d
t

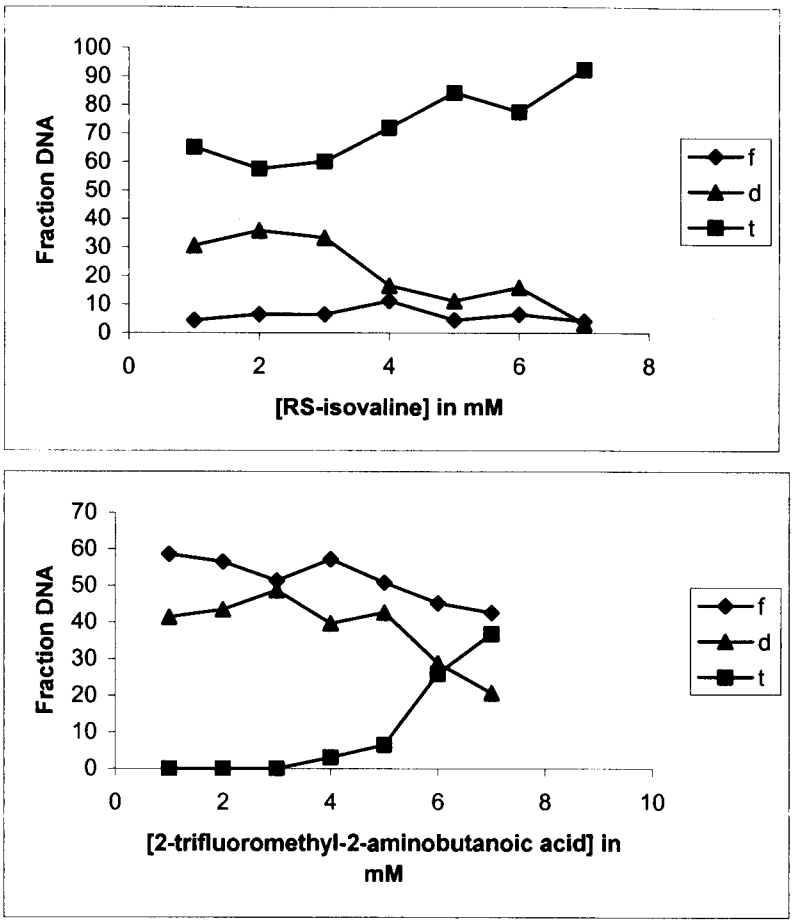


Figure 3-2B. Effects of different amino acids on free DNA and DgdR-DNA complexes in gel mobility shift assays. f is free DNA (fast-moving band in Figure 3-2A, d is the putative DgdR dimer-DNA complex (the middle electrophoresis band), and t is the putative DgdR tetramer-DNA complex (the slow-moving band in Figure 3-2A).

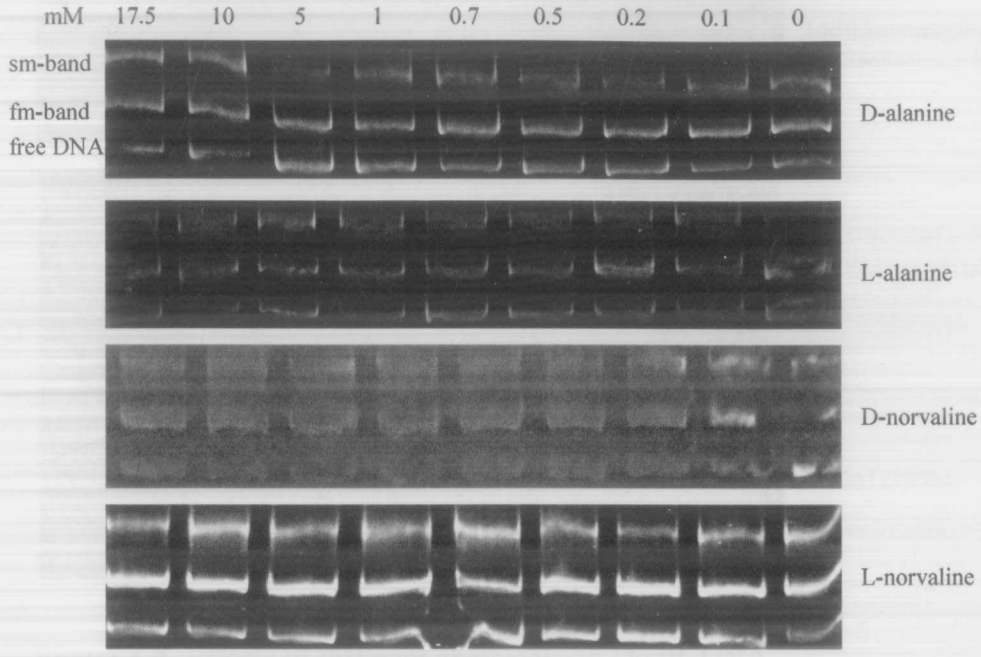


Figure 3-3. Effects of different amino acids on DgdR-DNA complex (II): Amino acids have no effects on DgdR-DNA complexes.

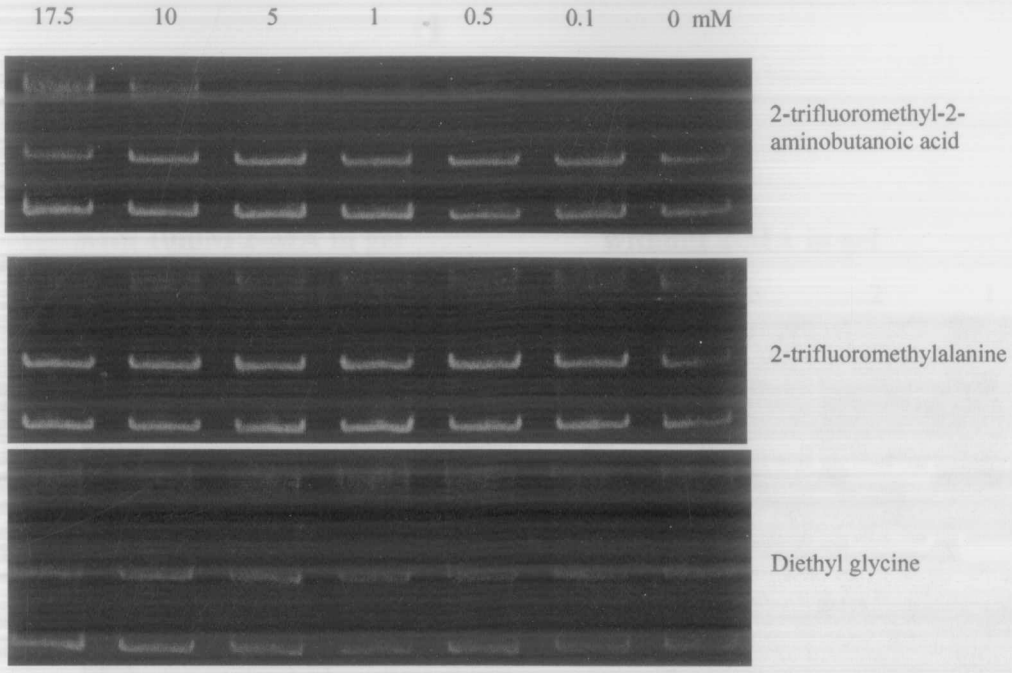


Figure 3-4. Effects of different chemicals on DgdR-DNA complex (III): Only 2-trifluoromethyl-2-aminobutanoic acid introduces minor changes on proportion and mobility of slow-moving bands at high concentrations.

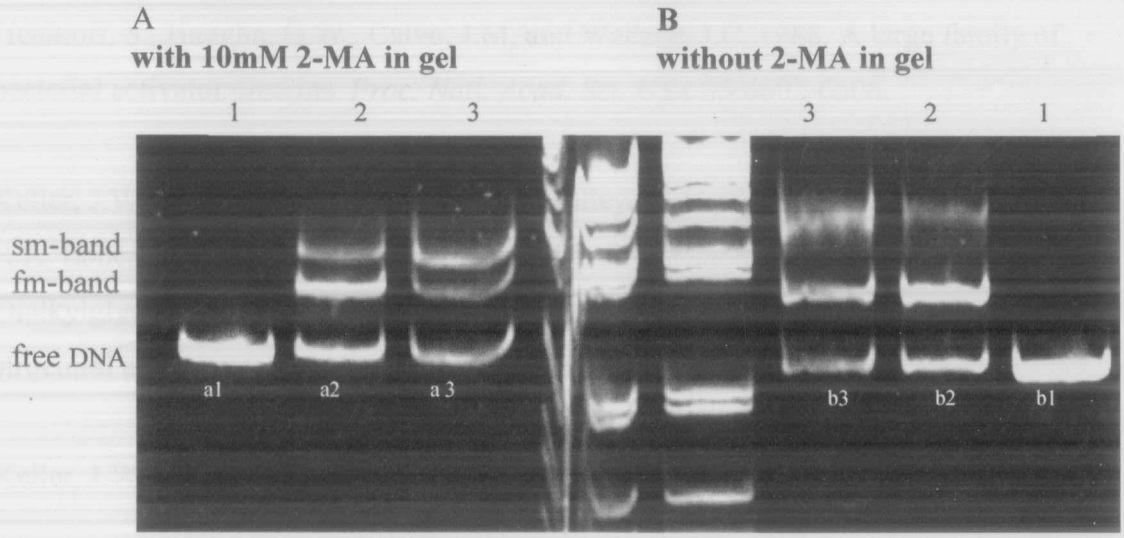


Figure 3-5. A Gel mobility shift assay with the inducer 2-methyalanine in gel (A) and not present in gel (B). sm-band: slow-moving band. fm-band: fast-moving band.

References:

Chugani, S.A., Parsek, M.R., and Chakrabarty, A.M. 1998. Transcriptional repression mediated by LysR-type regulator CatR bound at multiple binding sites. *J. Bacteriol.* 180 (9): 2367-2372.

Henikoff, S., Haughn, G.W., Calvo, J.M, and Wallace, J.C. 1988. A large family of bacterial activator proteins. *Proc. Natl. Acad. Sci. USA* 85:6602-6606.

Keller, J.W., Baurick, K.B., Rutt, G.C., O'Malley, M.V., Sonafrank, N.L., Reynolds, R.A, Ebbesson. L.O.E., and Vajdos, F.F. 1990. *Pseudomonas cepacia* 2,2-Dialkylglycine Decarboxylase. Sequence and expression in *Escherichia coli* of structural and repressor gene. *J. Biol. Chem.* 265(10): 5531-5539.

Keller, J.W. 1994 genbank.

Schell, M.A. 1993. Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* 47:597-626.

Allen-Daley, E., O'Brien, M.L., Bray-Hall, S.T., Stapleton, R.D., Chi, P., Sun, H. and Keller, J.W. Characterization of the *Burkholderia cepacia dgdR* gene encoding a LysR-type negative regulator of 2,2-dialkylglycine metabolism (in preparation).

Chapter 4

Construction of *dgdR* Fusion Gene: *dgdR/malE*

Abstract

DgdR, a LysR-type transcriptional regulator from *Burkholderia cepacia*, negatively regulates the expression of the *dgdA* gene, which encodes a dialkylglycine decarboxylase. The divergent *dgdR* and *dgdA* control the metabolism of 2-methylalanine in *B. cepacia*. The solubility of the DgdR protein, like that of most LysR-type proteins, is low. This common feature of the LysR protein caused the difficulty in the protein purification process. To enhance the solubility of this protein and facilitate future work on purification and characterization of the DgdR protein, the *dgdR* gene was fused to the *malE* gene, which encodes the maltose-binding protein (MBP). The *dgdR* gene was amplified from the plasmid pSB46 by polymerase chain reaction (PCR). The purified PCR fragment was digested by the restriction enzymes, *Eco*RI and *Ssp*I. The double-cut fragment was ligated into the plasmid vector pMAL-c2 that was linearized by *Eco*RI and *Xmn*I. The ligation mixture was transformed the competent host cells *E. coli* TB1. To screen for the presence of the insert, the transformants were collected for the restriction enzyme digestion analysis. Following this, automated DNA sequencing was conducted to further verify the presence of the *dgdR* gene.

Introduction

The DgdR protein and its divergent regulated gene, *dgdA*, control the metabolism of 2,2-dialkylglycines, such as 2-methylalanine, in *Burkholderia cepacia*.The DgdR protein and 2,2-dialkylglycine decarboxylase have been cloned and purified (Keller et al. 1990). The 2,2-dialkylglycine decarboxylase have been characterized by X-ray crystallography (Toney et al. 1995) and mechanistic studies have been conducted (Woon & Keller, in preparation). The nature of DgdR control has been characterized by the interaction of the purified DgdR protein with the *dgdA* promoter region, DNA footprinting analysis, and in vivo expression studies using the *xyl*E reporter gene constructs (Allen-Daley et al. in preparation). The results demonstrated that the DgdR protein negatively regulated the *dgdA* gene expression by blocking the entire region of the *dgdA* promoter.

A problem encountered during purification of DgdR protein was low solubility of this protein and other proteins in the LysR family (Schell 1993). This common feature complicates the purification process and precludes the possibility of crystallization. Recombinant DNA technology provides a method to solve this problem. In this study, the *dgdR* gene was cloned downstream and in frame with the *malE* gene of *E. coli*, whose product is the highly soluble maltose-binding protein (MBP). The existence of the insert gene was verified by automated DNA sequencing. However, further analyses and investigation were diverted to a new project, the study of the LysR homologs in hyperthermophilic archaeon. Studies on the fusion gene will be resumed at a later date.

The MBP fusion protein-expression system has been used to overexpress many heterologous proteins to high levels in *E. coli* (Davis et al. 1999, Kapust & Waugh 1999, Pryor & Leiting 1997). It had been reported that *E. coli* maltose-binding protein was uncommonly effective at promoting the solubility of polypeptides to which it is fused (Karust & Waugh 1999). A virulence plasmid regulatory protein from the LysR family, SpvR, had been fused to the maltose-binding protein (Grob & Guiney 1996). The MBP-SpvR fusion protein was successfully expressed in *E. coli* in large quantities. The fusion SpvR protein was able to induce the corresponding target gene expression in vivo and bind to the promoter region of its regulated cognate gene (Grob & Guiney 1996). In addition, it is possible to cut off the maltose-binding protein since a factor Xa recognition site has been engineerized just ahead of the insertion site of the foreign gene.

A MBP-DgdR fusion protein expressed from the vector produced in this study could be purified by a well-established method, affinity chromatography. MBP could be cut off by factor Xa (Marina et al. 1988) to generate the pure DgdR protein if it is necessary. If the MBP-DgdR fusion protein is soluble like the other MBP fusion protein, it will make it possible to crystallize this protein.

Materials and Methods

1. Plasmids, bacterial strains, and growth conditions

All plasmids and bacterial strains used in this study are compiled in Table 4-1. A circle map of plasmid vector pMAL-c2 is shown in Figure 4-1. All bacterial strains

were grown in Luria-Bertani (LB) liquid medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, H_2O,1 L) at 37°C with gentle shaking or on LB plates (10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g of agar, H_2O,1 L). When necessary, ampicillin was added to a final concentration of 80 μg/ml.

2. Primers and enzymes

The primers JK79 (5'-AGCTCGAAATATTTCAGTTAGAAAGGGGGCT-3') and JK80 (5'-CCACAGAATTCTCACAGCATCGGCAACAC -3') were used to amplify the *dgdR* gene from the plamid vector, pBS46. The *Ssp*I restriction enzyme cut site on primer JK79 and the *Eco*RI cut site on primer JK80 are underlined. Both primers were synthesized by Life Technologies. All the restriction enzymes, *Ssp*I, *Eco*RI, *Sal*I, *Hind*III, and *Xmn*I, used in this study were purchased from New England BioLabs. The *pfu* polymerase was purchased from Stratagene and T4 DNA ligase was purchased from Promega.

3. Plasmid isolation and purification

Plasmid isolation and purification follow steps described in Chapter 2 section 3 using Wizard™ Mini DNA purification kit.

4. Amplification of the *dgdR* gene from the pSB46 template

The Polymerase Chain Reaction (PCR) was used to amplify the entire *dgdR* gene from plasmid pBS46. The reaction was carried out in a 50 μl volume containing 1x *pfu*

polymerase buffer, 5 U of *pfu* polymerase, 200 μM 4 dNTPs, 1.5 mM $MgCl_2$, 1 μM of each primer, 30 ng of pBS46 as a template, and water. The reaction condition was 30 cycles of 94°C/45 s, 55°C/45 s, and 72°C/45 s on Coy TempCycler (Model 50/60, Coy Laboratory Products Inc). The PCR products were purified with StrataClean™ resin and Wizard™ PCR Preps DNA purification kit (Promega) as described in chapter 2.

5. Preparation of competent *E. coli* TB1 cells

Competent *E. coli* TB1 cells were prepared by the $CaCl_2$ method (Sambrook et al. 1987).

6. Construction of the *malE-dgdR* fusion gene: pMAL-c2/*dgdR*

6.1. Restriction enzyme digestion of the insert DNA fragment

The purified PCR product (the insert fragment) was digested by restriction enzymes, *Ssp*I and *Eco*RI to generate sticky ends. The digestion reaction was carried out as described in Chapter 2 starting with *Ssp*I. The 50 μl reaction contains 100-200 ng of the insert DNA, 1x *Ssp*I reaction buffer, 2 U of *Ssp*I, and water. The reaction mix was incubated at 37°C for 15-20 min. The completeness of *Ssp*I digestion was checked by 1% agarose electrophoresis. After the *Ssp*I digestion was complete, 2 U of the second restriction enzyme, *Eco*RI, was added to the reaction and continue to incubate at 37°C for another 15-20 min.

6.2. Restriction enzyme digestion of the vector pMAL-c2

The purified vector pMAL-c2 was digested by *Xmn*I and *Eco*RI, starting with *Xmn*I. The 50 µl reaction contains 100-200 ng of pMAL-c2, 1x *Xmn*I buffer, 2 U of *Xmn*I, and water. The reaction mixture was incubated at 37°C water-bath for 15-20 min. The *Xmn*I digestion was checked with 1% agarose gel electrophoresis. When the first enzyme digestion was complete, 2 U of the second restriction enzyme, *Eco*RI, was added and incubate at 37°C for another 15-20 min.

6.3. Purification of the digested insert fragment and vector pMAL-c2

The restriction enzyme digested PCR fragment and vector were purified as described in Chapter 2 with StrataClean™ resin (Stratagene) and Millipore ultrafree-MC filter (Millipore).

6.4. Ligation of the digested pMAL-c2 and the insert *dgd*R fragment and transformation of the competent *E. coli* TB1 cells

The purified insert fragment (*dgdR* fragment) and vector were subjected to a ligation reaction that was carried out in a 0.5 ml sterile Eppendorf tube. The 10 µl ligation reaction contains 100 ng of pMAL-c2 vector, about 60 ng of the *dgdR* fragment, 1x ligase buffer, 5 U of T_4 DNA ligase, and water. The mixture was incubated at room temperature overnight.

The mixture of the ligation reaction was used to transform the competent *E. coli* TB1 cells. Transformation was carried out by a standard transformation procedure (Sambrook et al. 1987).

7. Screen for fusion *dgdR* gene

The transformants growing on LB plates were collected and grown in 5 ml LB medium containing 80 μg/ml ampicillin for plasmid isolation and purification. Plasmid DNA was subjected to several restriction enzyme digestions (*Xmn*I and *Eco*RI, *Xmn*I and *Hind*III, and *Sal*I). The recombinant plasmid with the *dgdR* gene insert should generate two bands with sizes of 6,646 bp and 879 bp on an agarose gel after the *Xmn*I-plus-*Eco*RI digestion.

8. DNA sequencing

Automated DNA sequencing was conducted to verify the presence of the *dgdR* gene insert. Primer JK79 and JK80 were used to sequence the insert *dgdR* gene from both ends using pMAL-c2/*dgdR* as a template. The sequencing reaction consists of 400 ng of the purified recombined plasmid pBS46, 3 μl of Prism™ ready sequencing mix, 3 pmol of primer, and water to arrive a final volume of 15 μl. Cycle sequencing was carried out in 0.2 ml thin wall tubes on a Perkin-Elmer GeneAmp® PCR System 9600 thermal cycler. The reaction conditions were as follows: 1 cycle of 96°C, 1 min; 25 cycles of 96°C, 10 s; 50°C, 5 s; 60°C, 4 min. Excess nucleotides were removed by passing the products though Sephadex G-50 (Sigma) spin columns (Princeton Separations Centri-

Sep™). Samples were resolved on a 4.75% polyacrylamide gel by an ABI 373 DNA Sequencer. Sequence data were then analyzed with DNA Sequence Navigator and GenePro, 6.10.

Results

Two transformants grown on the LB agar plates were chosen for restriction enzyme and sequencing analysis. Plasmid DNA isolated from the transformants was digested by *Xmn*I and *Eco*RI. As shown in Figure 4-2, one of these two plasmids contained the insert *dgdR* gene that is 879 bp long.

Several additional restriction digestions were carried out to check the presence of the insert *dgdR* gene (Figure 4-3). When the plasmid pMAL-c2 without the presence of the *dgdR* insert was digested with *Eco*RI, only one band was observed as expected (lane 1, Figure 4-3). The recombinant plasmid, when cut with *Sal*I, should generate two fragments, 734 bp and 6791 bp, respectively (lane 2, Figure 4-3). When digested with *Hind*III and *Xmn*I, a band with 911 bp and a band with 6614 bp should be generated (lane 3 of Figure 4-3). The digestion with *Xmn*I and *Eco*RI generated a band with 879 bp that is the insert *dgdR* gene and a band with 6646 that is the pMAL-c2 plasmid. All the digestion results were the same as predicted in Figure 4-4.

The result from automated DNA sequencing demonstrated three point mutations shown in Table 4-2, the first one is located at position 458, T→C; the second one is at position 740, A→T; the third one at position 865, G→C, relative to the 5' end of the DgdR protein. The first mutation (T→C in CTC) generates proline instead of Leucine at

the corresponding place. The second mutation (A→T in GAT) replaces aspartic acid with valine. The third mutation (the first G→C in GTG) replaces valine with leucine. Since no mutations are located at the N-terminus of the *dgdR* gene, which is important for DNA recognition and binding, they should not affect the binding activity of this protein.

Discussion

The low solubility of DgdR protein raised difficulties in protein purification and future crystallization experiments as well. Therefore, a fusion gene was constructed to the *malE* gene from *E. coli* to improve the solubility. The *malE* gene encodes the maltose-binding protein that is highly soluble (Pryor & Leiting 1997). This property led to the protein fused with MBP expressed in large quantities and as a soluble protein (Grob & Guiney 1996, Chen & Gouaux 1996). At the same time, the MBP fusion protein demonstrated the similar binding activity with the wild-type protein (Grob & Guiney 1996). However, in this study, we have not induced the expression of *malE-dgdR* fusion gene since the project was diverted to search for soluble hyperthermophilic LysR homologs. It is possible in near future to induce the fusion gene expression by IPTG. The overexpressed gene products will be purified by a well-established affinity chromatography and the binding activity will be determined by an in vitro gel mobility shift assay. Moreover, the MBP part of the fusion protein can be easily cut off by Factor Xa. This will result in the pure DgdR protein. The binding

activities of the fusion protein and the pure DgdR protein could be compared by conducting gel mobility shift assays.

In conclusion, we have constructed the *dgdR* fusion gene with the *malE* gene from *E. coli* using DNA recombination techniques, and the structure was verified by restriction digestion and DNA sequencing. This construct will serve as a useful tool to study and understand the function of DgdR protein, crystallize DgdR protein, and facilitate the purification of this protein.

Table 4-1. Bacterial strains and plasmids

Strain or plasmid	Genotype, phenotype, or properties	Source
E. coli TB1	*araΔ(lac proAB)rpsL (Φ80 lacZΔM15) hsdR*	BioLabs
pMAL-c2	amp^r, pBR322 ori, P_{tac} promoter with the *malE* gene	BioLabs
pSB46	amp^r, pBTac1 with *dgdR* gene inserted downstream from the *tac* promoter	This Lab

Table 4-2. The *dgdR* DNA sequence comparison between wild-type and cloned *dgdR* gene

```
ATGCAAGGTAGAAAGGGGGCTAATACCTTGGGACGCTCGCTCGAAATCGACCTGCTGCGT 60
.............................................................

TCGTTCGTCGTGATCGCCGAGGTGCGCGCGCTCAGCGCGGCCGCGCGCGTCGGCCGGACG 120
.............................................................

CAGTCCGCGCTCAGCCAGCAGATGAAGCGGCTCGAGGATATCGTCGACCAGCCGCTGTTC 180
.............................................................

CAGCGCACCGGCCGCGGCGTGGTGCTGACGCACCCCGGCGAGCGGCTGCTCGTGCATGCG 240
.............................................................
.............................................................
CAGCGCATCCTGCGGCAGCACGACGAGGCAATGGCCGACCTGTGCGGCACGGGGTTGACG 300
.............................................................
.............................................................
GGGACGATCCGGTTCGGGTGCCCGGACGATTACGCGGAGGTGTTTCTGCCGCCGCTGCTG 360
.............................................................
.............................................................
CGGCAGTTTTCGAGCCAGCATCCGCAGGCGATCGTCGAAATCGTATGCGGGCCGACGCCG 420
.............................................................

CGGCTGCTCGAACAGCTCGAGAAGCGCGCGGTCGATCTCGCGATGATTTCATTGCCGGAC 480
.....................................c.......................
.....................................c.......................
GATGGGGCGAACGACGACATCATTCGTCGCGAGCAGCTGGTCTGGATCGGCTATCCGGGG 540
.............................................................
.............................................................
CTGGAGCCCGCGCATTTCGATCCGCTGCCGCTCGCGCTGTCCGATCCCGATACGCTCGAT 600
..................t
..................t..........................................
CACATCGCGGCCTGCGACGCGTTGCATCGCGCCGGTCGCGATTACCGCGTCGCGTATGCG 660

.............................................................
AGCAGCAGTCTCGCGGGGCTGATCGCGCTGGTGCGCTCGGGGCAGGCGTTCGCGGTGATG 720

.............................................................
ACGCAGACGGCCGTGCCGGCCGACCTGGCGATCGTCAACGGCGATCCGCGGTTGCCGCCG 780

.............................................................
TTGCCGGCGGTGGGCATTACGCTGAAGTTCGACCGGAAACGGCCGTCGCATCTGACGGCG 840

.............................................................
GCGTTCGCCGAGCATATTCGGGCCGTGTTGCCGATGCTG*                     900

.......................c
Dot indicates identical base. Lowercase letter indicates mutation.
Sequencing was done from both ends.
```

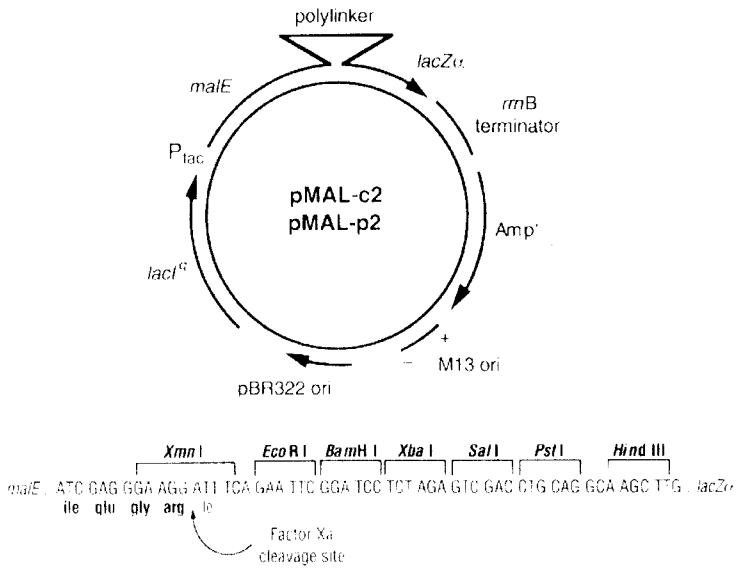


Figure 4-1. A map of pMAL-2 vector. The pMAL-c2 has 6646 base pairs. Arrows indicate the direction of transcription. Unique restriction enzyme sites are indicated (New England Biolabs instruction manual 1994).

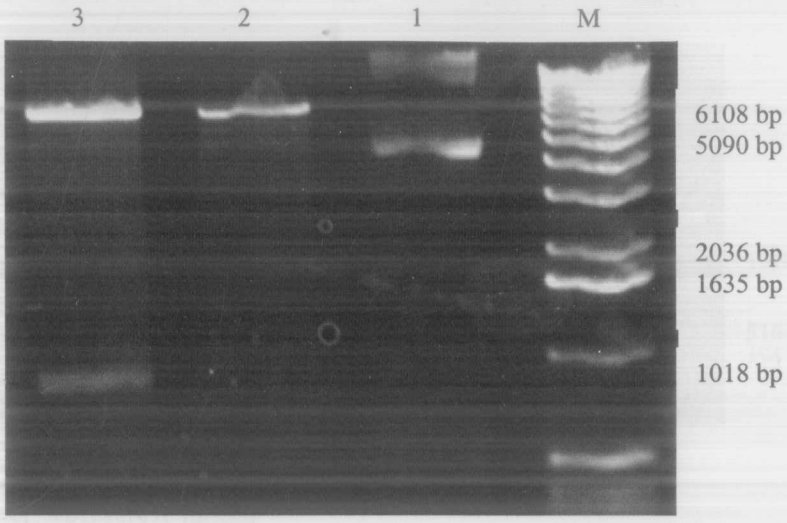


Figure 4-2. Recombinant pMAL-c2/*dgdR* from transformants.
M: 1kb DNA ladder
1: pMAL-c2 vector
2. pMAL-c2 digested by *Eco*RI
3. Recombinant pMAL-c2/*dgdR* digested by *Xmn*I and *Eco*RI.digested by *Xmn*I and *Eco*RI

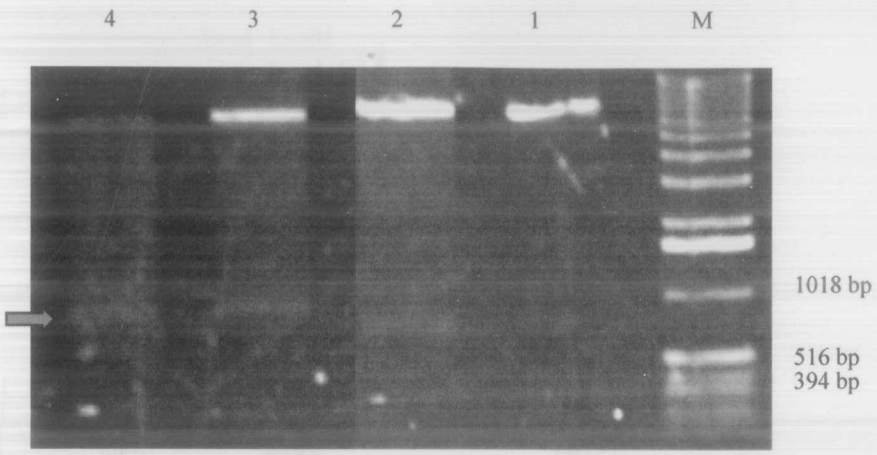


Figure 4-3. Restriction enzyme digestion analysis.
M: 1kb DNA ladder
1.pMAL-c2 digested by *Eco*RI (6646 bp)
2-4. Recombined pMAL-c2 digested by *Sal*I (734 bp and 6791 bp) *Hind*III+*Xmn*I (911 bp and 6614 bp), and *Xmn*I+*Eco*RI (879 bp and 6646 bp)
The arrow indicates *dgdR* insert gene.

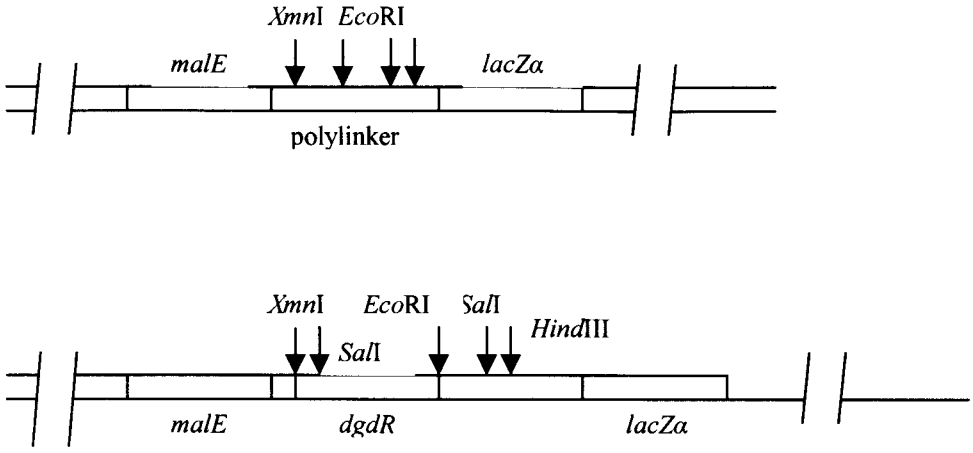


Figure 4-4. (top) Position of *malE* gene and polylinker on pMAL-c2 with *Xmn*I and *Eco*RI recognition sites labeled.
(bottom) Relative positions of restriction enzymes on the recombined plasmid: pMAL-c2 with the insert *dgdR* gene.

Reference:

Allen-Daley, E., O'Brien, M.L., Bray-Hall, S.T., Stapletion, R.D., Chi, P., Sun, H. and Keller, J.W. Characterization of the *Burkholderia cepacia dgdR* gene encoding a LysR-type negative regulator of 2,2-dialkylglycine metabolism (in preparation).

Chen, G. and Gouaux, J.E. 1996. Overexpression of bacterio-opsin in *Escherichia coli* as water-soluble fusion to maltose binding protein: efficient, regeneration of the fusion protein and selective cleavage with trypsin. *Protein Sci.* 5(3): 456-467

Grob, P. and Guiney, D.G. 1996. In vitro binding of the *Salmonella dubin* virulence plasmid regulatory protein SpvR to the promoter regions of *spvA* and *spvR*. *J. Bacteriol.* 178 (7): 1813-1820.

Kapust, R.B and Waugh, D.S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8(8): 1668-1674.

Keller, J.W., Baurick, K.B., Rutt, G.C., O'Malley, M.V., Sonafrank, N.L., Reynolds, R.A, Ebbesson. L.O.E., and Vajdos, F.F. 1990. *Pseudomonas cepacia* 2,2-dialkylglycine decarboxylase. Sequence and expression in *Escherichia coli* of structural and repressor genes. *J Biol. Chem.* 265(10): 5531-5539.

Maina, C.V., Riggs, P.D., Grandea, A.G. III Slatko, B.E., Moran, L.S., Tagliamonte, J.A., McReynolds, L.A. and Guan, C. 1988. A vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. *Gene* 74: 365-373.

Pryor, K.D. and Leiting, B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding protein double-affinity fusion system. *Protein Expr. Puri.* 10(3): 309-319.

Sachdev, D. and Chirgwin, J.M. 1998. Solubility of proteins from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expr. Puri.* 12(1): 122-132.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning: a laboratory manual. 2nd Edition. Cold Spring Harbor Laboratory Press.

Schell, M.A. 1993. Molecular biology of the LysR family of trnascriptional regulators. *Annu. Rev. Microbiol.* 47: 597-626.

Toney, M.D., Hohenester, E., Keller, J.W., and Jansonius, J.N. 1995. Structural and mechanistic analysis of two refined crystal structures of the pyridoxal phasphate-depedent enzyme dialkylglycine decarboxylase. *J. Mol. Biol.* 245: 151-179.

Woon, S. and Keller, J.K. Wild-type and mutant 2,2-dialkylglycine decarboxylases: the catalytic role of active site glutamine 52 investigated by site-directed mutagenesis and computer analysis. (in preparation)

Chapter 5

Recommendations for future research on MJ-LysR protein

Recommendation 1: hydroxyl (HO) radical and methylation interference (MI) footprinting to further characterize the MJ-LysR binding site.

To further characterize the MJ-LysR protein-binding site, hydroxyl (HO) radical and methylation interference (MI) footprinting can be employed. Combined with Dnase I footprinting data, these two methods can give information on close contacts between the MJ-LysR protein and the DNA backbone or bases. Dimethylsulfate (DMS) used in methylation interference footprinting methylates guanine residue at the N7 position in the DNA major groove and adenine residue at the N3 position in the minor groove. When the methylated DNA fragment was incubated with the protein, the guanine and adenine at the binding site were protected by the protein, thus cannot be cut by the cutting reagent (piperidine) in the following steps. Therefore, a contact region can be visualized as a blank region on an electrophoresis gel (Aububel et al. 1995). The hydroxyl (HO) radical footprinting gives the more detailed information on the protein binding region due to the more efficient cuts of hydroxyl radical along the backbone of the DNA (Dixon et al. 1991, Tullius et al. 1987). Hydroxyl radicals cleave DNA by abstracting hydrogen from C4 of the sugar in the minor groove. Protein binding over the binding site generally protects the sugar from cleavage. Therefore, the region protected by protein cannot be cut by hydroxyl radicals and will be visualized as a blank region on the gel (Carey & Smale 1999).

Recommendation 2: Dnase I footprinting to test the nature of MJ-LysR transcription regulation: how transcription factors TBP and TFB, and RNA polymerase affect the MJ-LysR mediated transcription.

As discussed above, in archaea, transcription factors, TBP and TFB, recruit RNA polymerase to the site to form a holoenzyme and stimulate the transcription. How the MJ-LysR affects this process remains unknown. It could block the assembly of one or more components of the transcriptional machinery: TBP, TFB, or RNA polymerase, or stimulate the assembly process. No matter what kind of role the MJ-LysR plays, Dnase I footprinting would give information on whether the TBP/TFB/DNA could form or not in the presence of the MJ-LysR protein. To do so, TBP/TFB and MJ-LysR could be incubated with the intergenic DNA fragment separately and the protein-DNA complexes formed then subjected to Dnase I digestion. The resulting footprint compared with the footprints obtained with TBP/TFB and MJ-LysR incubated with the same DNA fragment together will give information on how the MJ-LysR affects the binding of TBP/TFB. Gel mobility shift assays can be performed as well to determine whether MJ-LysR interferes with RNA polymerase recruitment or not. In gel mobility shift assays, different combinations of RNA polymerase, TBP/TFB, and MJ-LysR will incubate with the intergenic DNA region containing the putative promoter to detect how the MJ-LysR protein affects the formation of the holoenzyme. From the positions of the complexes formed on a nondenaturing polyacrylamide gel, the role that the MJ-LysR plays should be clarified.

Recommendation 3: In vitro transcription assay to test action mode of MJ-LysR protein on transcription: positively or negatively.

To investigate how the MJ-LysR protein influences the transcription, an in vitro transcription assay can be carried out as described by Hudelpohl (Hudelpohl et al. 1990, Darcy et al. 1999). Several archaeal TBP, TFB, and RNA polymerase are available commercially, such as those from *Sulfolobus acidocaldarius*. The different amount of the MJ-LysR protein will be added into each reaction system. From RNA products formed, the action mode of the MJ-LysR protein could be inferred.

Recommendation 4: A circular permutation analysis to calculate the bending angle formed by the binding of MJ-LysR protein on the intergenic DNA fragment.

DNA bending can be detected by a gel electrophoresis since DNA mobility is strongly dependent on the position of bending in the DNA molecules. The bending at the middle of the DNA fragment leads to the most retarded mobility while the bending at either end of the DNA fragment resulting the fastest band mobility (Wu & Crothers 1984). A circular permutation analysis is one of the methods employed most frequently to detect and calculate the bending angle induced by protein (Prenki et al. 1987, Wu & Crothers 1984). A DNA bending vector, pBend2, was constructed by a group of researchers (Kim et al. 1989). The double-stranded DNA fragment containing the binding site for the MJ-LysR protein can be inserted into the *Xba* I and *Sal* I sites in pBend2 as described by Pineiro (Pineiro et al. 1997). The resulting plasmid can be digested by a set of restriction enzymes to generate DNA fragments with identical

length but with the MJ-LysR binding site (the protein recognition sequence) located at variant positions along the intergenic DNA fragment. When the MJ-LysR protein incubates with these permuted DNA fragments, the mobility of the mixture detected by a polyacrylamide gel will be different, as the protein recognition sequence is located at different positions along the DNA fragment. The bending angle can be calculated using the empirical equation described by Thompson and Landy (1988) or the one modified by Elbright's laboratory (Zhou et al. 1993). The data from this experiment will provide direct information on bending induced by the MJ-LysR protein.

As mentioned above in Chapter 2, a piece of DNA segment occurs twice within the MJ-LysR binding site on the intergenic DNA fragment. What is the relationship between these two repeated segments? Is the binding of the MJ-LysR protein to each of them independent? It is possible to investigate the role of the two repeating segments within the MJ-LysR binding site using the same circular permutation system. A DNA fragment containing just one or both of these two repeating segments can be inserted into *Xba* I and *Sal* I site and incubated with MJ-LysR protein. The migrations of mixtures can be detected on a nondenatured polyacrylamide gel, thus the bending angle can be calculated and the pattern of new bands can be detected. By comparing with the bending angles and the band patterns obtained, the role of these repeating segments and the relationship between them should be elucidated.

Recommendation 5: site-directed mutagenesis to study the key bases in the MJ-LysR recognition.

Site-directed mutagenesis has been used to analyze the critical nucleotides in the interaction of a protein with its target promoter region (Parsek et al. 1994, Hayes & Tullius 1989, Cho & Winans 1993). It is a well-established technique to create the desired mutations at certain positions. Nucleotides within the binding site obtained by the DNase I footprinting will be target for site-directed mutagenesis. The site-directed mutagenesis kits are commercially available to generate mutations of the intergenic DNA fragment. The mutations will be checked by automated DNA sequencing. The intergenic DNA fragment with the targeted mutation will be incubated with the MJ-LysR protein followed by a gel mobility shift assay to check the binding affinity of the MJ-LysR protein to the mutated DNA fragments. The importance of nucleotides in the intergenic DNA fragment for the binding of MJ-LysR protein could be estimated by this way.

Recommendation 6: Computational modelling of MJ-LysR protein.

Although the best answer to get structural information of a protein comes from a crystal structure, computational methods can provide some important information when the crystal structure is not available. To date, several online free softwares, such as DINAMO (Bentz et al. 1999) and ProMod and Swiss-Model (Peitsch 1996, Schwede et al. 2000) are available and could be used to build a model for a protein whose 3-D structure remains unknown. However, in order to get a reliable model, the similarity

between the target sequence, which is the sequence of protein with the unknown structure, and the template sequence, which is the sequence of the protein with known 3-D structure used as a reference to build the model, has to be greater than 25%. In the LysR family, the only protein with the crystal structure being solved is CysB from *Klebsiella aerogenes*. The homology between MJ-LysR and CysB is only 18%. Such a low homology makes it impossible to use those softwares mentioned above to build a reliable model. However, it is still possible to use other molecular modeling software such as *Sybyl*, *HyperChem* to build a model. This can be done in a stepwise fashion. The first thing to do is to get the best alignment between the MJ-LysR protein and CysB, then the amino acids in CysB that are different from those in MJ-LysR will be mutated to the corresponding amino acids in the MJ-LysR to lead to a new structure, at last the energy will be minimized using the build-in tool in the software. Nevertheless, the reliability of the model should be subject to the further investigation.

References:

Aubulbel, F.M., Brent, R., Kinston, R.E., Moore, D.D., Smith, J.A., and Struhl, K. 1995. Current Protocols in Molecular Biology (volume II). John Wiley & Sons Inc.

Bentz, J., Baucom, A., Hansen, M., and Gregoret, L.M. 1999. *DINAMO*: interactive protein alignment and model building. *Bioinform.* 15(4): 309-316

Carey, M. and Smale, S.T. 1999. Transcriptional regulation in eukaryotes: concepts, strategies, and techniques. *Cold Spring Harbor Laboratory Press.*

Cho, K. and Winans, S.C. 1993. Altered-function mutations in the *Agrobacterium tumefaciens* OccR protein and in an OccR-regulated promoter. *J. Bacteriol.* 175(23): 7715-7719

Darcy, T.J., Hausner, W., Awery, D.E., Edwards, A.M., Thomm, M., and Reeve, J.N. 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription in vitro. *J. Bacteriol.*181(14): 4424-4429

Dixon, W.J., Hayes, J.J., Levin, J.R., Weidner, M.F., Dombroski, B.A., and Tullius, T.D. 1991. Hydroxyl radical footprinting. *Methods Enzymol.* 208: 380-413

Hayes, J.J. and Tullius, T.D. 1989. The missing nucleoside experiment: a new technique to study recognition of DNA by protein. *Biochem.* 28(24): 9521-9527

Hudepohl, U., Reiter, W., and Zillig, W. 1990. In vitro transcription of two rRNA genes of the archaeobacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific initiation. *Proc. Nat. Acad. Sci. USA* 87: 5851-5855

Kim, J., Zwieb, C., Wu, C., and Adhya, S. 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* 85: 15-23

Parsek, M.R, Ye, R.W., Pun, P., and Chakrabarty, A.M. 1994. Critical nucleotides in the interaction of a LysR-type regulator with its target promoter region. *J. Biol. Chem.* 259(15): 11279-11284

Peistch, M.C. 1996. ProMod and Swiss-Model: internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* 24(1): 274-279.

Pineiro, S., Olekhnovich, I., and Gussin, G.N. 1997. DNA bending by the TrpI protein of *Psedomonas aeruginosa. J. Bacteriol.* 179(17): 5407-5413

Prentki, P., Pham, M.H., and Galas, D.J. 1987. Plasmid permutation vectors to monitor DNA bending. *Nucleic Acids Res.* 15:10060

Schwede, T., Diemand, A., Guex, N., and Peistch, M.C. 2000. Protein structure computing in the genomic era. *Res. Microbiol.* 151(2): 107-112.

Thompson, J.F. and Landy, A. 1988. Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucleic Acids Res.* 16(20): 9687-9705

Tullius, T.D., Bombroski, B.A., Churchill, M.E.A., and Kam, L. 1987. Hydroxyl redical footprinting: a high-resolution method for mapping protein-DNA contacts. *Methods enzymol.* 155: 537-558

Wu, H.M. and Crothers, D.M. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature* 308: 509-513

Zhou, Y., Zhang, X., and Ebright, R.H. 1993. Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proc. Natl. Acad. Sci. USA* 90: 6081-6085